Grant Agreement No.: 732569
Call: H2020-ICT-2016-2017
Topic: ICT-13-2016
Type of action: CSA

HUB4NGI

D2.2 NGI GUIDE V2

Revision: v.1.7

Work package	WP 2
Task	Tasks 2.1, 2.2
Due date	31/12/2017
Submission date	20/12/2017
Deliverable lead	IT Innovation
Version	1.8
Authors	Steve Taylor, Brian Pickering, Michael Boniface (IT Innovation)
Reviewers	Alexander Mikroyannidis (OU), John Delaney (IDC)

Abstract	This deliverable summarises the current work concerning three related topics following on from D2.1: gap analysis to determine subject areas for further consultation; methods, practice and initial results from two consultations in different subject areas; and how researchers and innovators can be supported to create beneficial and effective solutions to real world applications in the NGI.
Keywords	NGI, evidence

DISCLAIMER

The information, documentation and figures available in this deliverable are written by the "A Collaborative Platform to Unlock the Value of Next Generation Internet Experimentation" (HUB4NGI) project's consortium under EC grant agreement 732569 and do not necessarily reflect the views of the European Commission.

The European Commission is not liable for any use that may be made of the information contained herein.

COPYRIGHT NOTICE

Project co-funded by the European Commission in the H2020 Programme		
Nature of the deliverable:	R	
Dissemination Level		
PU	Public, fully open, e.g. web	✓
CL	Classified, information as referred to in Commission Decision 2001/844/EC	
CO	Confidential to HUB4NGI project and Commission Services	

* R: Document, report (excluding the periodic and final reports)

DEM: Demonstrator, pilot, prototype, plan designs

DEC: Websites, patents filing, press & media actions, videos, etc.

OTHER: Software, technical diagram, etc.

EXECUTIVE SUMMARY

This deliverable summarises the current work concerning three related topics following on from D2.1: gap analysis to determine subject areas for further consultation; methods, practice and initial results from two consultations in different subject areas; and how researchers and innovators can be supported to create beneficial and effective solutions to real world applications in the NGI.

Following a gap analysis, two consultation subject areas have been selected: “Responsible Autonomous Machines” and “Echo Chambers and Fake News”. The main reasoning for this selection is that they were seen as important, with significant Research & Development & Innovation (R&D&I) potential, but were not yet addressed in detail within the current version of the planned work programme.

- Responsible Autonomous Machines are typically autonomous algorithms or applications of Artificial Intelligence (AI) whose actions need to be explainable and governed from both a legal and ethical standpoint, because they are either safety-critical or impact the lives of citizens in significant ways. The consultation concerns investigation into responsibility aspects, societal impacts and risks of AI and autonomous machines.
- Echo Chambers are situations where citizens may not be receiving full, accurate or unbiased information via their interactions with the Internet. Especially important is content verification (e.g. combatting fake news) and alternative strategies to the current, where content providers profile Internet users and deliver customised and filtered news, content feeds or search results to users that may present a biased perspective.

The current status of the consultations at the time of writing (December 2017) is that expert participants have been identified and recruited for the Responsible Autonomous Machines consultation. The consultation has been launched, initial responses received, analysed and presented back to the participants for comment. The consultation is due to finish in Q1 2018, and results will be made available (most likely as a white paper) as soon as they are ready.

The Echo Chambers consultation is in the process of identification of experts, and it is planned that the consultation be launched with consultees in January 2018. The expected timescale for this consultation is in the order of three months, so results can be expected in Q2 2018.

To support innovators, this deliverable has provided an initial investigation into innovation pathways. To avoid ambiguity, a clear distinction is made between the Innovation Process and its result, the Innovation itself. The Innovation and its Process has been contextualised within its broader ecosystem and is broken down into five main components:

1. The *Ambition* is the overall vision, what needs to be achieved; and defines the rationale and motivation behind a given innovation. Note that the ambition may be influenced both by society as a whole, what it expects and what it will not accept, as well as by stakeholder perceptions and expectations. This feeds into:
2. The *Innovation Process* itself where an idea is evaluated and implemented or elaborated to produce a recognisable outcome. As well as responding to an ambition, the innovation process is informed and constrained by two constructs:
3. On the one hand, there are *Stakeholders* who have an interest in whatever the outcome of the process may be, but also in how it is achieved. On the other hand, *Knowledge* in broad and general terms will constrain what can be done and influence the choices on how it can be done.
4. Finally, *Society* is the main beneficiary of the innovation outcome, but may also constrain it (via knowledge and stakeholders) or seed innovation (via ambition).

Co-funded by the Horizon 2020
Framework Programme of the European Union

The innovation model will be evaluated through prototyping within WP3, informed by the issues highlighted in the two consultations. Interaction with WP1 is planned, to revisit the KPIs in the light of the issues raised via the consultations.

TABLE OF CONTENTS

LIST OF FIGURES

Co-funded by the Horizon 2020
Framework Programme of the European Union

ABBREVIATIONS

AI	Artificial Intelligence
AR	Augmented Reality
EC	European Commission
FAANG	Facebook, Amazon, Apple, Netflix, Google
GAFA	Google, Amazon, Facebook, Apple
GDPR	General Data Protection Regulation
IoT	Internet of Things
IP	Internet Protocol
NGI	Next Generation Internet
OECD	Organisation for Economic Co-operation and Development
ORD	Open Research Data
RAM	Responsible Autonomous Machines
SDN	Software Defined Networking
TCP	Transmission Control Protocol
VR	Virtual Reality

1 INTRODUCTION

This deliverable covers work from T2.1 - NGI Vision and Strategy, and T2.2 - NGI Innovation Pathways and Gap Analysis. It reports on two major pieces of work in progress: consultations to determine recommendations for upcoming NGI work programmes, and a breakdown of the elements and process for innovation within the NGI.

Mapping this onto tasks, the consultations contribute to T2.1 because their outcome contains recommendations for future research and innovation in the NGI work programmes. The consultations also contribute to the gap analysis of T2.2 because they concentrate on areas that are seen as important by the community but not significantly addressed in the current drafts of upcoming work programmes. The work on NGI innovation elements and process contributes to T2.2's "innovation pathways" element.

The deliverable first concentrates on the consultations, followed by innovation support. The methodology for the consultations is described next, including the reasoning for selection of the particular subject areas via a gap analysis. This is followed by the status, preparation of the consultations, their execution and the available interim results for the consultations. Following this, innovation configuration, pathways and support are discussed. To avoid ambiguity, the term "innovation" is defined with reference to multiple external sources, and a domain model of the innovation elements and process in the context of its environment of external actors is provided. The components of the domain model are then broken down further to illustrate the entities and concerns.

This deliverable is very much a snapshot of a number of works-in-progress. Interim results where available are presented, but are indicative only. The full results and methods towards the results will be presented in D2.3. This is not due until the end of the NUB4NGI project, so as soon as results are ready, they will be made available as white papers so as not to delay their effectiveness.

Co-funded by the Horizon 2020
Framework Programme of the European Union

2 CONSULTATIONS

2.1 INTRODUCTION

This section describes the ongoing consultations conducted by IT Innovation for WP2. As a result of the work done in Q1 & Q2 of 2017 and reported in D2.1, detailed consultations are being executed by HUB4NGI on subjects highlighted in the synthesis reported in D2.1. The aim of the consultations is to provide details on technical subjects that are identified as important but are not yet obviously covered in the work programme planning to date. The chosen method to provide these details is consultation with worldwide experts using a structured methodology - the Delphi Method[1]. Figure 1 below shows the methodology diagrammatically.

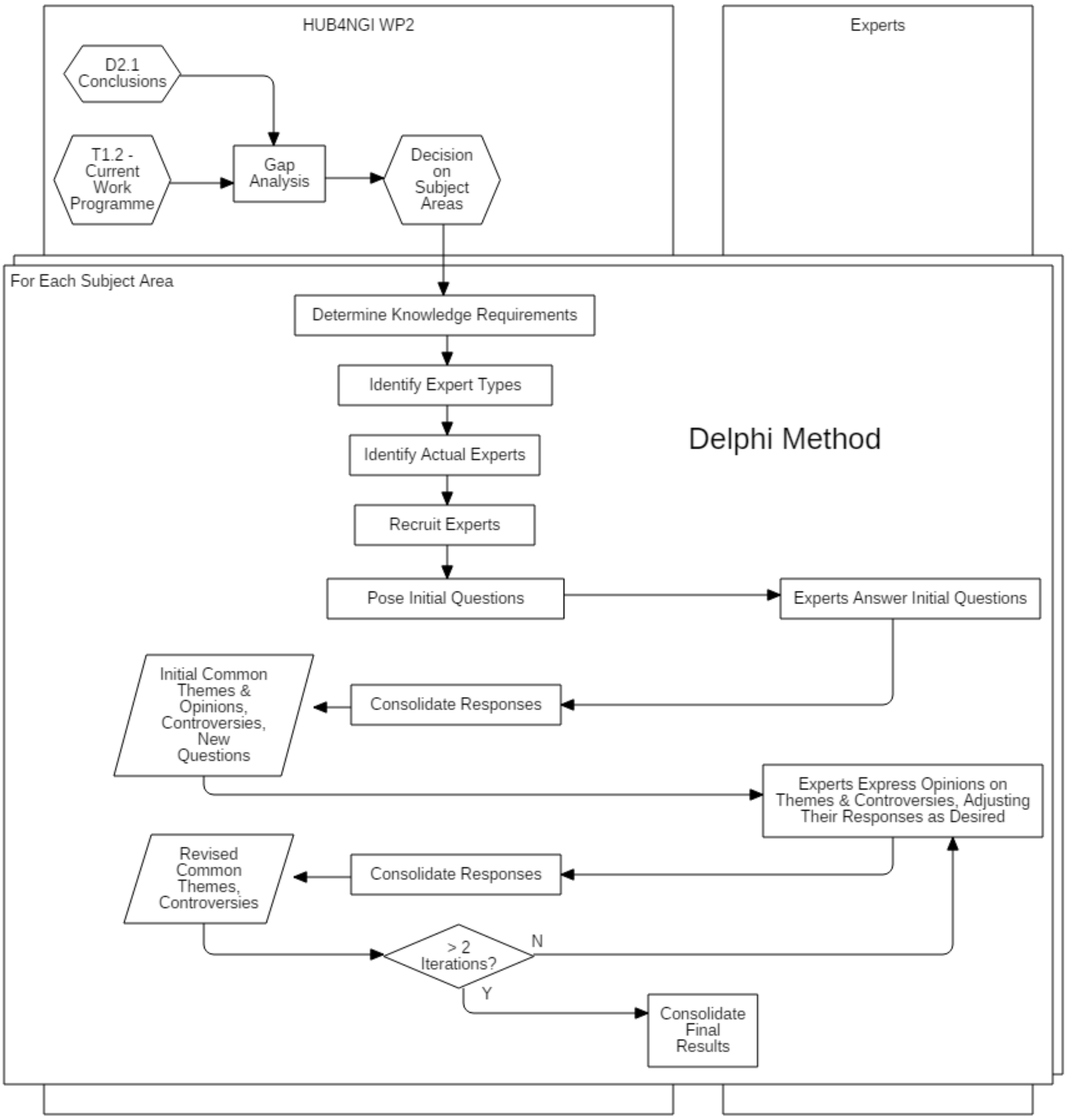

HUB4NGI WP2
Experts
D2.1 Conclusions
T1.2 - Current Work Programme
Gap Analysis
Decision on Subject Areas
For Each Subject Area
Determine Knowledge Requirements
Identify Expert Types
Identify Actual Experts
Recruit Experts
Delphi Method
Pose Initial Questions
Experts Answer Initial Questions
Initial Common Themes & Opinions, Controversies, New Questions
Consolidate Responses
Experts Express Opinions on Themes & Controversies, Adjusting Their Responses as Desired
Revised Common Themes, Controversies
Consolidate Responses
> 2 Iterations?
N
Y
Consolidate Final Results

FIGURE 1: CONSULTATION METHODOLOGY

These main elements of the methodology are covered in detail in the next sections, followed by details of each consultation.

[1] See, for example, https://thepsychologist.bps.org.uk/volume-22/edition-7/delphi-method ; and https://www.rand.org/content/dam/rand/pubs/papers/2008/P3558.pdf

Co-funded by the Horizon 2020
Framework Programme of the European Union

2.2 GAP ANALYSIS FOR SUBJECT AREAS

D2.1 provided a synthesis of nine previous external consultation exercises conducted in the recent past on NGI, covering where the consultations agreed and where they differed. A total of nine major theme clusters were identified across these previous consultations. The theme clusters are presented below, annotated with pros and cons regarding subsequent detailed consultation potential to inform technical research, development and innovation within the NGI.

Area	Pros	Cons
Decentralised Control	Clear societal need	Technical R&D not clear
Decentralised Infrastructures	Clear technical R&D Rich, varied subject area	Already known (in ICT24)
Privacy Enhancement	Clear societal need Clear technical R&D Rich, varied subject area	Already known (in ICT24)
Responsible Autonomous Machines	Clear societal need Potential for multidisciplinary R&D	
Innovation Networks & Multidisciplinary Design	Supporting innovation	Technical R&D not clear Already supported by T2.3, Innovation Pathways
Legislation Process	Clear need Multidisciplinary opportunity	Technical R&D not clear
Echo Chambers & Fake News	Hot topic (fake news) Potential for technical R&D	Need to determine exact technical R&D ➔ Consultation
Economics & Wealth Distribution	Clear need	Technical R&D not clear Probably out of ICT scope
Trust & Security	Clear need Clear technical R&D	Already well known (although not in ICT 24)

Out of the subjects identified in the D2.1 analysis, two stood out as being seen as important, with significant R&D&I potential, but were not yet addressed in detail within the planned work programme. These are “Responsible Autonomous Machines” and “Echo Chambers & Fake News”, and these have been selected for more detailed consultations within HUB4NGI.

- Responsible Autonomous Machines are typically autonomous algorithms or applications of Artificial Intelligence (AI) whose actions need to be explainable[2] and governed from both a legal and ethical standpoint, because they are either safety-critical or impact the lives of citizens in significant ways. The consultation concerns investigation into responsibility aspects, societal impacts and risks of AI and autonomous machines. There is clearly a need for multidisciplinary collaborative research into the ethical, legal, and societal impact of AI and autonomous machines, and into how they can be regulated and certified for compliance to safety and ethical standards.
- Echo Chambers are situations where citizens may not be receiving full, accurate or unbiased information from their interactions with the Internet. Especially important is content verification (e.g. combatting fake news) and alternative strategies to the current, where content providers profile Internet users and deliver customised and filtered news, content feeds or search results to users that may present a biased perspective. There is currently an online survey targeted at citizens and journalists / organisations regarding fake news run by the EC[3,4], and the intention is that our planned consultation complements this with a slightly wider perspective involving the additional but related elements discussed above.

2.3 CONSULTATION STATUS

The current status of the consultations at the time of writing (December 2017) is that expert participants have been identified and recruited for the Responsible Autonomous Machines consultation. The consultation has been launched, and initial responses have been received, analysed, and presented back to the participants for comment. The consultation is due to finish in Q1 2018, and results will be made available (as most likely a white paper) as soon as they are ready.

The Echo Chambers consultation is in the process of identification of experts. It is likely that the actual consultation process will begin in January 2018; if it were to begin in December 2017, there is a risk that momentum will be lost over the Christmas holiday in the crucial first round of consultation. The expected timescale for this consultation is in the order of three months, so results can be expected in Q2 2018.

2.4 CONSULTATION METHODOLOGY

2.4.1 Choice of Methodology

The consultations' chief aim is to provide concrete, justifiable and credible recommendations for the planning of upcoming NGI work programmes. Therefore, the recommendations should be based on consensus amongst informed opinions. It may not be possible to arrive at consensus for all areas, and this is a valid result in itself: controversies can highlight areas for future research to clarify them or to provide more evidence.

Public surveys were considered as a consultation mechanism, meaning anyone could participate, but were not seen as attractive because, owing to the self-selecting nature of the respondent population, it is not clear that the respondents are knowledgeable enough about the consultation's subject matter, whether the population has inherent bias or is impartial, or whether the population covers enough of the interdisciplinary expertise needed. In addition, a number of

[2] The term "explainable" acknowledges the "Explainable AI" (XAI) movement in Artificial Intelligence. See e.g. https://en.wikipedia.org/wiki/Explainable_Artificial_Intelligence. The movement is based on the need to address the problem of "black box AI" where it is not clear to a human observer how or why the decisions of AI systems came about.

[3] https://ec.europa.eu/info/consultations/public-consultation-fake-news-and-online-disinformation_en

[4] https://ec.europa.eu/digital-single-market/en/fake-news

Co-funded by the Horizon 2020
Framework Programme of the European Union

large-scale general Internet consultations involving the general public exist already, and were synthesised in the work reported in D2.1.

Given the specific nature of the subject areas, it was considered that targeted consultations were preferred, where experts in relevant fields could be selected based on their reputations and consulted in a managed way. This provided reassurance about the validity of the opinions expressed, through the experts' track records and reputations in their respective fields.

The chosen methodology for the consultation is the Delphi Method[5], a well-established pattern that aims to determine consensus or highlight differences from a panel of selected consultees. These properties make the Delphi Method ideally suited for the purposes of targeted consultations with experts with the intention of identifying consensuses for recommendations.

Because the consultations target world-level experts, naturally busy people, it is necessary that the consultations be as low time-cost and as non-intrusive as possible for the experts. An additional property of the Delphi Method is that it is necessarily anonymous during its runtime - participants do not know the identities of other participants. The major justification for this property is to avoid participants being influenced by the name and reputation of other participants and bandwagon-jumping to agree with them. This has the useful consequence that the consultations can be run remotely without the need for travel to meetings or booking time slots for teleconferences etc, and so there is little intrusion into the experts' time. The actual mechanism is that of an Internet-based survey to which the targeted experts are invited. This is remote and non-interactive so the experts can fill in the survey at a time that suits them.

2.4.2 Delphi Method - Description

The Delphi Method arrives at consensus by iterative rounds of consultations with the expert panel. Initial statements made by participants are collated with other participants' statements and presented back to the panel for discussion and agreement / disagreement. This process happens over a number of rounds, with subsequent rounds refining the previous round's statements based on feedback from the panel so that a consensus is reached, or controversies highlighted.

For the consultations described here, experts are asked to participate in a remote, non-interactive, anonymous consultation that consists of three iterations, with consolidation of the answers in between iterations. This consultation is administrated by a facilitator (Steve Taylor) who manages the consultation process. Each round is run as a separate online survey, and the format of the rounds are described as follows.

- Round 1. A selected panel of experts are invited to participate in Round 1 based on their reputation in a field relevant to the core subject of this consultation. Round 1 is a web survey containing a background briefing note to set the scene, accompanied by two broad, open-ended questions to which participants can make any free-form text responses they wish[6].
- Round 2. Using standard qualitative techniques such as thematic analysis, the collected corpus of responses from Round 1 are independently encoded to generate assertions. The assertions are presented back to the participants, who are given an opportunity to confirm or revise their opinions in the light of the consolidated previous results. This uses a structured format web survey (e.g. the participants can agree or disagree with the assertions on a 4-point Likert scale).

[5] Linstone, H.A. and Turoff, M. eds., 1975. The Delphi method: Techniques and applications (Vol. 29). Reading, MA: Addison-Wesley.

[6] https://thepsychologist.bps.org.uk/volume-22/edition-7/delphi-method

Co-funded by the Horizon 2020
Framework Programme of the European Union

- Round 3. The results of Round 2 are collated, refining the consensuses and disagreements, which are presented back to the participants who can confirm or refine their opinions further, again using a structured format web survey.

The results of the third round are collated to determine the final consensus and disagreements, which form the output recommendations of the consultation.

2.4.3 Expert Selection

Expert selection is a critical part of the Delphi Method, because clearly the experts determine the results of the consultation. Criteria observed in the selection of experts for the consultations are discussed as follows.

The experts must form a multidisciplinary panel covering relevant subject areas. For each consultation, a "knowledge requirements" exercise is needed to determine the different subjects where knowledge is required for the consultation.

A good target for the number of experts in the panel is 10-20. If the panel contains less than 10 experts, it is not likely that enough coverage of relevant subjects will be possible. If the panel contains more than 20 experts, it is likely to be difficult to manage by the facilitator.

The effort to determine which experts to invite must not be underestimated – it is a time-consuming task to determine the knowledge requirements, identify candidate experts within the subject areas and assess their level of expertise and suitability. Clearly, the level of expertise is important, and world-level experts should be targeted. It must therefore be acknowledged that experts are busy people and so it is reasonable to expect a low response rate to invitations to participate. We are assuming a 10-20% response rate, so in order to achieve the desired expert numbers of 10-20 experts in the panel, we must invite 80-100 experts.

2.4.4 Ethical approval

The consultations are studies involving external participants, so by the regulations of the University of Southampton, ethical approval must be applied for and granted before the consultations can go ahead.

The full application for the first consultation, Responsible Autonomous Machines, is given in Section 5 as an appendix. The application consists of four documents:

- An application form, describing the study and evaluating the risks (in this case the risks are negligible).
- A Participant Information Sheet, given to potential participants and describing the nature of the study and what will happen during the course of the study.
- A consent form, to be filled in by participants to indicate their consent to take part in the study.
- A Data Protection Plan, describing what personal data will be collected from the participants, and what will be done with it.

The bulk of these documents can be repurposed for the application for ethical approval for the consultation on Echo Chambers, as the format of the study, the personal data collected and its processing are identical to the Responsible Autonomous Machines study.

2.4.5 Collation of Input

Once responses are returned from experts, they need to be collated so as to provide input for the next round or to determine the final results of the consultation.

For the first round of consultation, the responses are in textual format. They need to be collated into thematic groups that indicate the general subject areas of R&D&I. Within the themes, assertion statements are needed based on the initial responses. These form the basis of the second round, where they are assessed by the experts. Qualitative thematic analysis should be performed independently by two analysts, and the results compared to confirm the major subject themes, the interpretations of the source texts and their transformation into assertion statements.

The second and third rounds are scored using Likert scales that describe agreement or disagreement with the assertion statements from the previous round (e.g. 1 = strongly disagree to 4 = strongly agree). Analysis of these results is quantitative, and consensus can be determined by statistical methods such as mean (showing the most common position on the scale) and standard deviation (showing the spread of respondents' votes on the scale).

2.5 RESPONSIBLE AUTONOMOUS MACHINES

The justification for the subject of Responsible Autonomous Machines is given in D2.1[7] and the NGI Emerging Research Challenges White Paper[8]. This work also provides starting point themes for further investigation.

> *The so-called "responsible machines" are typically autonomous applications of AI whose actions need to be regulated because they are either safety-critical or impact the lives of citizens in significant ways, such that regulation is needed. Autonomous vehicles are an exemplary case.*
>
> *There is a pressing need for research and discussion involving multidisciplinary teams from the legal, sociological and technical domains to provide answers to ethical and legal questions surrounding responsible machines. Key questions include the following, and research is needed to address them.*
>
> *The issue of legal and moral responsibility for AI systems is a critical unresolved question. Who or what takes responsibility for an AI system's decisions or actions, especially if an AI system causes harm? Could it ever be the case that an AI system is a legal entity and bears responsibility for its actions in its own right?*
>
> *There is currently a debate regarding the application of ethics to responsible machines. Some advocate that ethics should be designed into AI technology, while others argue that it is the application of the AI technology that needs ethical governance. Investigation into the pros and cons of each argument is needed. Related to this issue is the question of how AI should be regulated. Should there be design regulations for "ethical AI", or should the applications of AI be regulated?*
>
> *Transparency of AI decision making is a key aspect of the so-called "algorithmic accountability". There are fears amongst experts that AI decisions may deliberately or inadvertently include bias or discrimination. Investigation is needed into how the algorithms can explain their decisions, and how bias or discrimination can be avoided.*
>
> *Responsible machines often operate in safety-critical modes, where their actions or inactions can cause harm to humans. Safety-critical software needs commitments from developers to provide updates to fix bugs and security*

[7] Steve Taylor & Michael Boniface, HBU4NGI D2.1 NGI GUIDE V1

[8] Taylor, S., Boniface M. Next Generation Internet: The Emerging Research Challenges - Key Issues Arising from Multiple Consultations Concerning the Next Generation of the Internet. https://ec.europa.eu/futurium/en/next-generation-internet/next-generation-internet-emerging-research-challenges-key-issues-arising

> *flaws, and there is an open question on how commitments can be acquired from creators of AI technology to issue patches for safety-critical flaws over the long term, including what will happen should a safety-critical AI developer go out of business.*

2.5.1 Knowledge Requirements & Expert Selection

Directly from the text above, based on previous analysis reported in D2.1, the following major themes are present:

- Legislation & Regulation of AI
- Ethics of AI
- Legal and Moral Responsibility of AI
- Explainable and Transparent AI

Using these themes as a basis, an exploratory literature survey was conducted to determine related subject fields of expertise and candidate experts' names needed for the Responsible Autonomous Machines consultation. Each of the themes above was explored using standard tools and methods such as Google Scholar, Microsoft Academic, standard Google searches and following links from Wikipedia pages to gain a background into the theme, related themes, as well as influential people contributing work within the theme. As a result of this exploration, the list of themes expanded to include:

- Algorithmic Accountability (transparency)
- Machine Ethics
- Bias & discrimination in AI systems
- Philosophy & computation
- Psychology of computation
- Sociology & the social impact of AI (human-machine networks)
- Robotics & responsibility ("killer robots")
- AI risks & threats ("existential threats" & AI Safety)
- Superintelligence
- Artificial Moral Agency
- Economics & computation
- Epistemology
- Law & computation

Some of these themes are connections between two subjects (notably a social science and computation), while others are exemplary handles that have been adopted by communities of like-minded researchers to give a name to their field (e.g. Algorithmic Accountability and Machine Ethics). These are often the result of seminal works that spawn the field or a new direction, and clearly the authors of these seminal works should be targeted as experts.

To find experts, a number of methods was used, and these are described briefly below.

- Google Scholar searches were made for the key subject theme areas, looking for seminal or highly cited publications and their authors.
- The proceedings from key recent conferences in relevant subjects were examined for publications and authors.

Co-funded by the Horizon 2020
Framework Programme of the European Union

- Attendee lists of previous Internet consultations were examined for possible candidate experts.
- General Google searches for the themes resulted in web pages of many different kinds. Relevant institutes were found, and their staff pages provided a source of possible experts. Other websites were interest pages of e.g. academics, or describing a relevant initiative.
- HUB4NGI partners were consulted for suggestions of experts, and OU responded with suggestions.

The many searches provided cross-correlation and reinforcement of relevant experts, for example the same person was found in multiple related searches, and major contributors to a particular field became apparent.

The result of these investigations was a spreadsheet describing names of experts, their contact details, with notes on their specialisms. The experts' names will not be disclosed[9], but a total of 88 experts roughly evenly distributed between the subject themes above were invited to the consultation.

The literature survey also provided input material for the background briefing note (next), which provided context on the subject matter by discussing starting point themes and issues of the consultation. This was distributed to all experts invited to the consultation.

2.5.2 Background

This section contains the text for a briefing note that was circulated to potential participants to set the background context of the Responsible Autonomous Machines consultation.

RESPONSIBLE AUTONOMOUS MACHINES – CONSULTATION WITH EXPERTS:
Background & Gateway Questions

This briefing note aims to set the context for a consultation with domain experts from multiple disciplines about important research questions, topics and themes in and around the subject area of "*Responsible Autonomous Machines*". As a definition, ***Responsible Autonomous Machines are typically autonomous algorithms or applications of AI whose actions need to be explainable and governed from both a legal and ethical standpoint because they are either safety-critical or impact the lives of citizens in significant ways.*** The themes of the consultation strongly correlate with and support those of the current Beneficial AI movement[10].

The purpose of the consultation is to determine a research agenda that will inform the European Commission on the important research topics surrounding Responsible Autonomous Machines, and thus assist them to determine a future work programme of research within the H2020 framework and FP9.

The overall context of the problem domain is given in the Background, next, to indicate starting-point themes, but consultees are encouraged to suggest any additional related themes as they see fit.

As AI and automated systems have come of age in recent years, they promise ever more powerful decision making, providing huge potential benefits to humankind through their performance of mundane, yet sometimes safety-critical tasks, which they can perform better

[9] Part of the consent for the consultation are questions whether the expert wishes to be named as an author of a publication of the results of the consultation. Those that consented will be named in the publication, otherwise any experts approached or consulted will be kept strictly anonymous.

[10] https://futureoflife.org/bai-2017/

than humans[11,12]. Research and development in these areas will not abate and functional progress is unstoppable, but there is a clear need for ethical considerations applied to[13,14] and regulatory governance of[15,16], these systems, as well as AI safety in general[17] with well-publicised concerns over the responsibility and decision-making of autonomous vehicles[18] as well as privacy threats, potential prejudice or discriminatory behaviours of endemic web applications[19,20,21,22]. Influential figures such as Elon Musk[23] and Stephen Hawking[24] have voiced concerns over the potential threats of undisciplined AI, with Musk describing AI as an existential threat to human civilisation and calling for its regulation. Recent studies into the next generation of the Internet such as Overton[25] and Takahashi[26] concur that regulation and ethical governance of AI and automation are necessary, especially in safety-critical systems and critical infrastructures.

Over the last decade, machine ethics has been a focus of increased research interest[27]. Citing the well-known no-win situation of a runaway trolley[28], Anderson & Anderson identify issues around increasing AI enablement not only in technical terms[29], but significantly in the societal context of human expectations and technology acceptance transplanting the human being making the ethical choice with an autonomous system[30]. Anderson & Anderson also describe different mechanisms for reasoning over machine ethics[30]. Some of these concern the encoding of general principles (e.g. principles following the pattern of Kant's categorical imperatives[31]) or domain-specific ethical principles, while others concern the selection of precedent cases of ethical decisions in similar situations (e.g. SIROCCO[32]) and a further class considers the consequences of the action under question (act utilitarianism – see Brown[33]). An open research question concerns which mechanism, or which combination of mechanisms, is appropriate. Anderson & Anderson advocate a hybrid approach following the pattern of Ross's *prima facie duties*[34] (duties "at first sight") where basic principles may be adjusted as necessary via case-

[11] Donath, Judith. The Cultural Significance of Artificial Intelligence. 14 December 2016. https://www.huffingtonpost.com/quora/the-cultural-significance_b_13631574.html

[12] Ruocco, Katie. Artificial Intelligence: The Advantages and Disadvantages. 6th February 2017. https://www.arrkgroup.com/thought-leadership/artificial-intelligence-the-advantages-and-disadvantages/

[13] Bostrom, N. & Yudowsky, E. (2014). The ethics of artificial intelligence. In Ramsey, W. & Frankish, K. (eds) *The Cambridge handbook of artificial intelligence*, 316-334.

[14] https://www.wired.com/story/ai-research-is-in-desperate-need-of-an-ethical-watchdog/

[15] Scherer, Matthew U., Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies (May 30, 2015). Harvard Journal of Law & Technology, Vol. 29, No. 2, Spring 2016. Available at SSRN: https://ssrn.com/abstract=2609777 or http://dx.doi.org/10.2139/ssrn.2609777

[16] Vincent C. Müller (2017) Legal vs. ethical obligations – a comment on the EPSRC's principles for robotics, Connection Science, 29:2, 137-141, DOI: 10.1080/09540091.2016.1276516

[17] https://futureoflife.org/2017/09/21/safety-principle/

[18] Bonnefon, J-F, Shariff, A. & Rahwan, I (2016). The social dilemma of autonomous vehicles. *Science* 352(6293), 1573-1576.

[19] http://www.independent.co.uk/news/world/americas/facebook-rules-violence-threats-nudity-censorship-privacy-leaked-guardian-a7748296.html

[20] http://www.takethislollipop.com/

[21] https://www.youtube.com/watch?v=4obWARnZeAs

[22] Crawford, K. (2016) "Artificial intelligence's white guy problem." The New York Times (2016).

[23] Musk, E. (2017) Regulate AI to combat 'existential threat' before it's too late. *The Guardian,* 17th July, 2017

[24] Stephen Hawking warns artificial intelligence could end mankind, BBC News, 2 December 2014. http://www.bbc.co.uk/news/technology-30290540

[25] DAVID OVERTON, NEXT GENERATION INTERNET INITIATIVE – CONSULTATION - FINAL REPORT MARCH 2017 https://ec.europa.eu/futurium/en/content/final-report-next-generation-Internet-consultation

[26] Takahashi, Makoto. Policy Workshop Report Next Generation Internet - Centre for Science and Policy Cambridge Computer Laboratory. Centre for Science and Policy (CSaP) in collaboration with the Cambridge Computer Laboratory. 1-2 March 2017. https://ec.europa.eu/futurium/en/system/files/ged/report_of_the_csap_policy_workshop_on_next_generation_Internet.docx Retrieved 2017-06-19.

[27] Allen, C., Wallach, W., & Smit, I. (2006) Why machine ethics? *IEEE Intelligent Systems, 21(4),* 12-17

[28] Thomson, J. J. (1985). The trolley problem. *The Yale Law Journal*, *94*(6), 1395-1415.

[29] Anderson, M., & Anderson, S. L. (Eds.). (2011). *Machine ethics*. Cambridge University Press.

[30] Anderson, Michael, and Susan Leigh Anderson. "Machine ethics: Creating an ethical intelligent agent." AI Magazine 28, no. 4 (2007): 15. https://doi.org/10.1609/aimag.v28i4.2065

[31] https://plato.stanford.edu/entries/kant-moral/

[32] McLaren, Bruce M. "Extensionally defining principles and cases in ethics: An AI model." Artificial Intelligence 150, no. 1-2 (2003): 145-181. https://doi.org/10.1016/S0004-3702(03)00135-8

[33] Brown, Donald G. "Mill's Act-Utilitarianism." The Philosophical Quarterly 24, no. 94 (1974): 67-68.

[34] Ross, William David. The right and the good. Oxford University Press, 1930, new edition 2002.

Co-funded by the Horizon 2020
Framework Programme of the European Union

based induction of a course of action, often modifying basic principles in the light of a new situation by reference to previous experience.

A long-debated key question is that of legal and moral responsibility of autonomous systems. Who or what takes responsibility for an autonomous system's actions? Calverley[35] considers the question from a legal perspective, asking whether a non-biological entity can be regarded as a legal person. If a non-biological entity such as a corporation can be regarded as a legal person, then why not an AI system? The question then becomes one of intentionality of the AI system and whether legal systems incorporating penalty and enforcement can provide sufficient incentive to AI systems to behave within the law. Matthias[36] poses the question whether the designers of an AI system can be held responsible for the system they create, if the AI system learns from its experiences, and is therefore able to make judgements beyond the imagination of its designers. Beck[37] discusses the challenges of ascribing legal personhood to decision-making machines, arguing that society's perceptions of automata will need to change should a new class of legal entity appear. In addition, careful consideration will be needed regarding the eligibility conditions for membership of this new class.

Transparency of autonomous systems is also of concern, especially given the opaque (black-box) and non-deterministic nature of AI systems such as Neural Networks. The so-called discipline of "explainable AI" is not new: in 2004, Van Lent et al[38] described an architecture for explainable AI within a military context and in 2012, Lomas et al[39] demonstrated a system that allows a robot to explain its actions by answering "why did you do that?" types of question. More recently, in response to fears of accountability for automated and AI systems, the field of *algorithmic accountability reporting* has arisen *"… as a mechanism for elucidating and articulating the power structures, biases, and influences that computational artefacts exercise in society"*[40]. In the USA, the importance of AI transparency is clearly identified, with DARPA recently proposing a work programme for research towards explainable AI (XAI)[41,42].

The above issues and others are encapsulated in the "Asilomar AI Principles" [43], a unifying set of principles that are widely supported and should guide the development of beneficial AI, but how should these principles be translated into a research agenda for the EC?

Following are starting point questions, aimed at kicking off the consultation process. They are by nature open-ended, so as to give respondents maximum freedom in their input.

- *What research is needed to address the issues that Beneficial AI and Responsible Autonomous Machines raise?*
- *Why is the recommended research important?*

2.5.3 Status

The current status at the time of writing is that Round 2 of the Responsible Autonomous Machines consultation is underway.

[35] Calverley, D.J., 2008. Imagining a non-biological machine as a legal person. Ai & Society, 22(4), pp.523-537.

[36] Matthias, A., 2004. The responsibility gap: Ascribing responsibility for the actions of learning automata. Ethics and information technology, 6(3), pp.175-183.

[37] Beck, S., 2016. The problem of ascribing legal responsibility in the case of robotics. AI & society, 31(4), pp.473-481.

[38] Van Lent, Michael, William Fisher, and Michael Mancuso. "An explainable artificial intelligence system for small-unit tactical behavior." In Proceedings of the National Conference on Artificial Intelligence, pp. 900-907. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2004.

[39] Lomas, Meghann, Robert Chevalier, Ernest Vincent Cross II, Robert Christopher Garrett, John Hoare, and Michael Kopack. "Explaining robot actions." In Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction, pp. 187-188. ACM, 2012. https://doi.org/10.1145/2157689.2157748

[40] Diakopoulos, N., 2015. Algorithmic accountability: Journalistic investigation of computational power structures. Digital Journalism, 3(3), pp.398-415.

[41] DARPA 2016 - Broad Agency Announcement - Explainable Artificial Intelligence (XAI) DARPA-BAA-16-53, August 10, 2016. https://www.darpa.mil/attachments/DARPA-BAA-16-53.pdf

[42] Gunning, David. "Explainable artificial intelligence (xai)." Defense Advanced Research Projects Agency (DARPA), nd Web (2017).

[43] The Asilomar AI Principles, proposed during the Beneficial AI 2017 Conference, Asilomar, California, 5-8 January 2017. https://futureoflife.org/ai-principles/

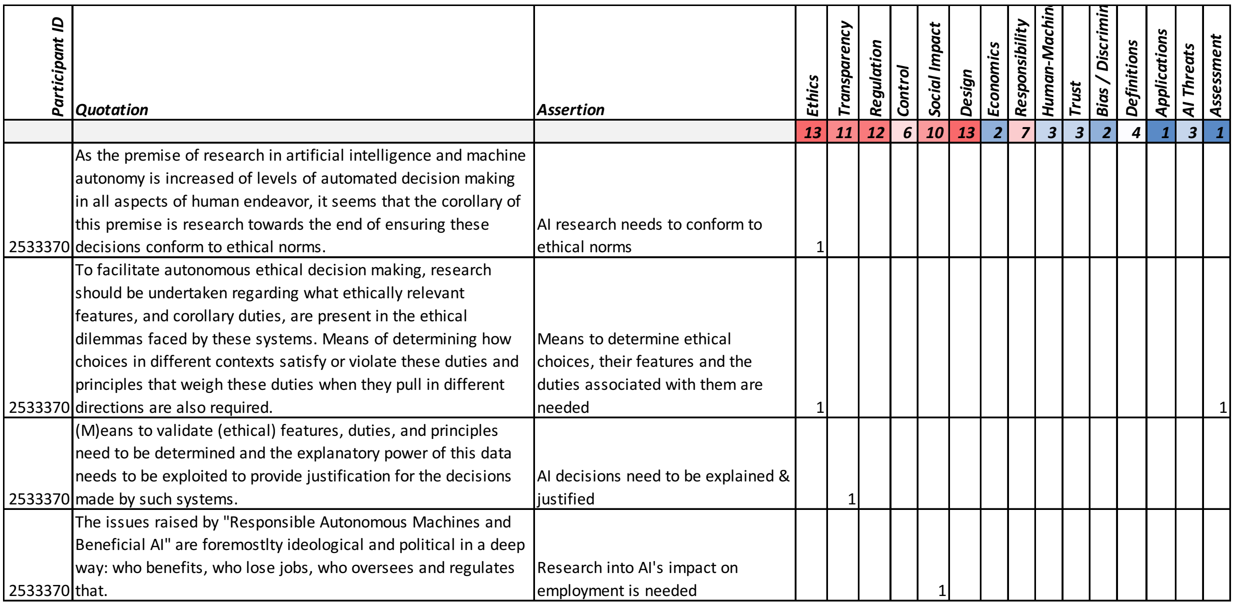
Ethical approval was applied for and granted for the consultation, as it contains negligible risks for participants. The full ethical approval application documents are given in the Appendix in Section 5, and much of the material will be repurposed for the Echo Chambers consultation, as the format and risks of the two consultations are the same.

After the initial knowledge requirements exercise, 88 experts were targeted and these were invited to participate in Round 1 of the consultation. A total of 14 responded – 16%, typical for unsolicited consultations, of which 12 provided detailed textual answers to the two gateway questions, which was collated to provide assertion statements used in Round 2.

2.5.4 Round 1 Analysis

The aim of the analysis was to determine assertion statements from Round 1 responses that could be used as input for Round 2, where they are presented to the participants for agreement / disagreement and comments, with a view to establishing consensus regarding each assertion statement.

Each respondent's textual answers were scrutinised for opinions, statements and recommendations. Where one was found, the relevant quotation from the text was recorded along with a summary to form a draft assertion. Each quotation was also classified according to its broad thematic subject matter, and as new thematic classes appeared from observation of the textual responses, they were added to the classifier list. The purpose of the thematic classifiers was to determine subject groupings for presentation of assertions in Round 2. An example of early analysis is shown in Figure 2. A summary assertion associated and the classification of subject matter can be seen for each relevant source quotation.



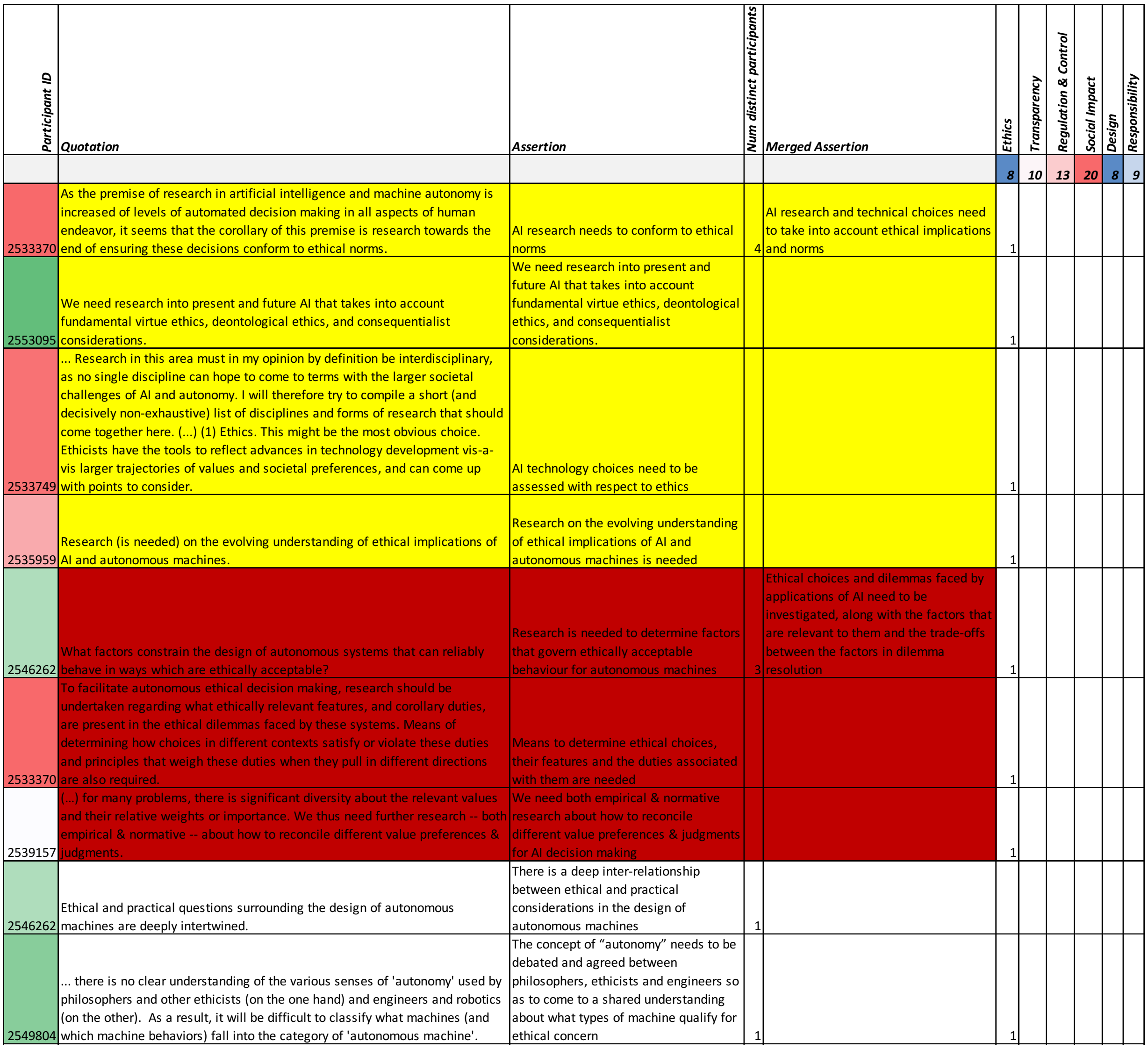
Participant ID	Quotation	Assertion	Ethics	Transparency	Regulation	Control	Social Impact	Design	Economics	Responsibility	Human-Machin	Trust	Bias / Discrimin	Definitions	Applications	AI Threats	Assessment
			13	11	12	6	10	13	2	7	3	3	2	4	1	3	1
2533370	As the premise of research in artificial intelligence and machine autonomy is increased of levels of automated decision making in all aspects of human endeavor, it seems that the corollary of this premise is research towards the end of ensuring these decisions conform to ethical norms.	AI research needs to conform to ethical norms	1														
2533370	To facilitate autonomous ethical decision making, research should be undertaken regarding what ethically relevant features, and corollary duties, are present in the ethical dilemmas faced by these systems. Means of determining how choices in different contexts satisfy or violate these duties and principles that weigh these duties when they pull in different directions are also required.	Means to determine ethical choices, their features and the duties associated with them are needed	1														1
2533370	(M)eans to validate (ethical) features, duties, and principles need to be determined and the explanatory power of this data needs to be exploited to provide justification for the decisions made by such systems.	AI decisions need to be explained & justified		1													
2533370	The issues raised by "Responsible Autonomous Machines and Beneficial AI" are foremostlty ideological and political in a deep way: who benefits, who lose jobs, who oversees and regulates that.	Research into AI's impact on employment is needed					1										

FIGURE 2: ASSERTION SUMMARISATION & SUBJECT CLASSIFICIATION

After analysis of all the respondents' textual answers, it became apparent that multiple participants were expressing the same opinion about a subject (albeit worded differently). It also became apparent that there were too many classification subjects.

The next phase of analysis clustered multiple quotations that expressed the same opinion into a single summary assertion, whilst recording how many participants expressed that opinion. It also collapsed the subject categories into a smaller number, so as to provide a reasonably low number of subject headings for Round 2. An example of this analysis is shown in Figure 3.

Co-funded by the Horizon 2020
Framework Programme of the European Union



Participant ID	Quotation	Assertion	Num distinct participants	Merged Assertion	Ethics	Transparency	Regulation & Control	Social Impact	Design	Responsibility
					8	10	13	20	8	9
2533370	As the premise of research in artificial intelligence and machine autonomy is increased of levels of automated decision making in all aspects of human endeavor, it seems that the corollary of this premise is research towards the end of ensuring these decisions conform to ethical norms.	AI research needs to conform to ethical norms	4	AI research and technical choices need to take into account ethical implications and norms	1					
2553095	We need research into present and future AI that takes into account fundamental virtue ethics, deontological ethics, and consequentialist considerations.	We need research into present and future AI that takes into account fundamental virtue ethics, deontological ethics, and consequentialist considerations.			1					
2533749	... Research in this area must in my opinion by definition be interdisciplinary, as no single discipline can hope to come to terms with the larger societal challenges of AI and autonomy. I will therefore try to compile a short (and decisively non-exhaustive) list of disciplines and forms of research that should come together here. (...) (1) Ethics. This might be the most obvious choice. Ethicists have the tools to reflect advances in technology development vis-a-vis larger trajectories of values and societal preferences, and can come up with points to consider.	AI technology choices need to be assessed with respect to ethics			1					
2535959	Research (is needed) on the evolving understanding of ethical implications of AI and autonomous machines.	Research on the evolving understanding of ethical implications of AI and autonomous machines is needed			1					
2546262	What factors constrain the design of autonomous systems that can reliably behave in ways which are ethically acceptable?	Research is needed to determine factors that govern ethically acceptable behaviour for autonomous machines	3	Ethical choices and dilemmas faced by applications of AI need to be investigated, along with the factors that are relevant to them and the trade-offs between the factors in dilemma resolution	1					
2533370	To facilitate autonomous ethical decision making, research should be undertaken regarding what ethically relevant features, and corollary duties, are present in the ethical dilemmas faced by these systems. Means of determining how choices in different contexts satisfy or violate these duties and principles that weigh these duties when they pull in different directions are also required.	Means to determine ethical choices, their features and the duties associated with them are needed			1					
2539157	(...) for many problems, there is significant diversity about the relevant values and their relative weights or importance. We thus need further research -- both empirical & normative -- about how to reconcile different value preferences & judgments.	We need both empirical & normative research about how to reconcile different value preferences & judgments for AI decision making			1					
2546262	Ethical and practical questions surrounding the design of autonomous machines are deeply intertwined.	There is a deep inter-relationship between ethical and practical considerations in the design of autonomous machines	1							
2549804	... there is no clear understanding of the various senses of 'autonomy' used by philosophers and other ethicists (on the one hand) and engineers and robotics (on the other). As a result, it will be difficult to classify what machines (and which machine behaviors) fall into the category of 'autonomous machine'.	The concept of "autonomy" needs to be debated and agreed between philosophers, ethicists and engineers so as to come to a shared understanding about what types of machine qualify for ethical concern	1		1					

FIGURE 3: ASSERTION CLUSTERING AND SUBJECT COLLAPSE

The figure shows a total of four output assertions. The first two (yellow and red in the figure) are sets of quotations expressing similar statements or opinions, and they have been combined into a single merged assertion. The number of distinct participants in a merged assertion set (i.e. the number of different people expressing the same opinion) is also shown. The two white assertions are cases where only one person made a statement or expressed a particular opinion, so there is no need to merge them with other assertions and in these cases the original summary assertion is used in the output.

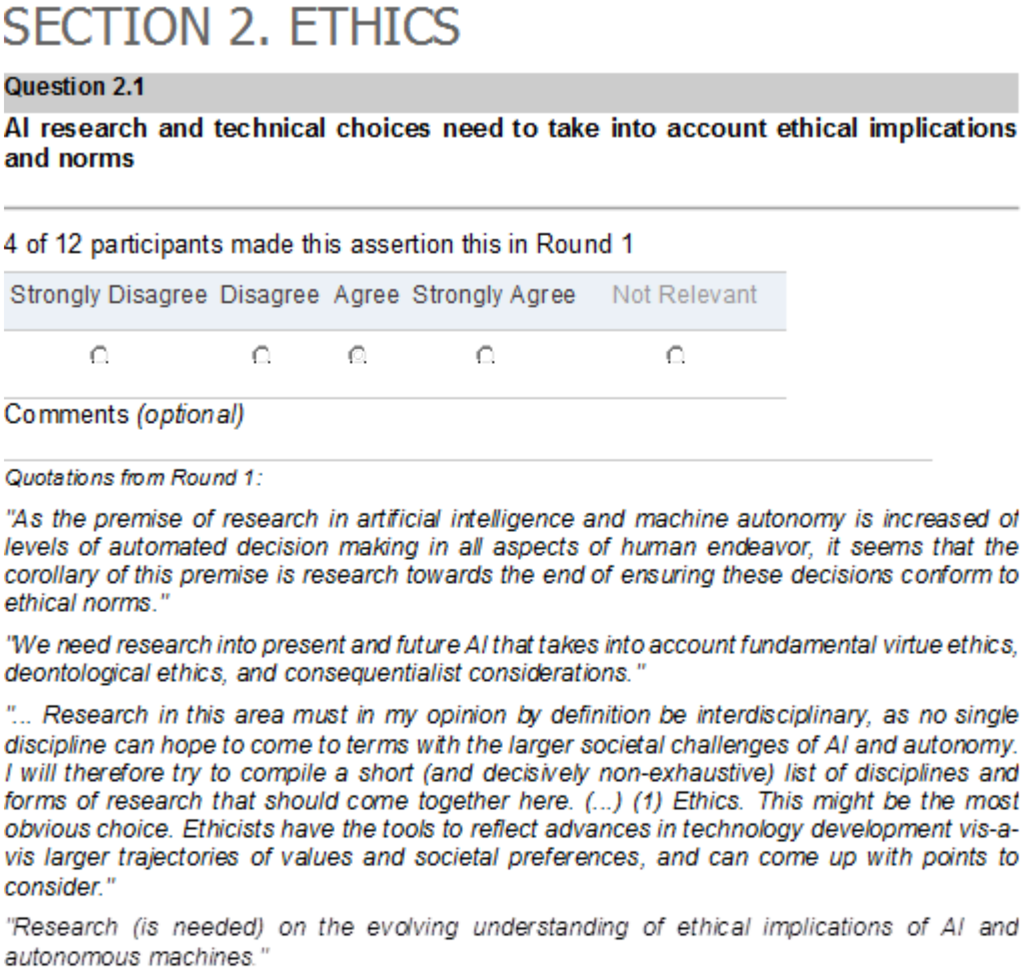
The assertions are grouped into the following subject category sections:

- *Ethics* (ethical implications for AI & autonomous machines and their applications),
- *Transparency* (considerations regarding transparency, justification and explicability of AI & autonomous machines' decisions and actions),
- *Regulation & Control* (regulatory aspects such as law, and how AI & automated systems' behaviour may be monitored and if necessary corrected or stopped),
- *Social Impact* (how society is impacted by AI & autonomous machines),
- *Design* (design-time considerations for AI & autonomous machines) and
- *Responsibility* (issues and considerations regarding moral and legal responsibility for scenarios involving AI & autonomous machines).

Co-funded by the Horizon 2020
Framework Programme of the European Union

A standard thematic analysis methodology (TA) was adopted for the free-form text that respondents had provided. As an entirely inductive approach, it was felt that TA would provide a solid starting point to identify issues which could be used to generate assertions for the next round of the consultation. Two researchers coded the original text independently. The overlap of interim themes was good (4 out of 6 themes were clearly the same). The researchers then met to discuss and agree the final set of themes. No formal analysis of inter-coder reliability[44] was therefore felt necessary.

Figure 4 shows the presentation of an output assertion in Round 2, as seen by the consultation participants for agreement / disagreement and comment[45].



SECTION 2. ETHICS

Question 2.1

AI research and technical choices need to take into account ethical implications and norms

4 of 12 participants made this assertion this in Round 1

Strongly Disagree	Disagree	Agree	Strongly Agree	Not Relevant
○	○	○	○	○

Comments *(optional)*

Quotations from Round 1:

"As the premise of research in artificial intelligence and machine autonomy is increased of levels of automated decision making in all aspects of human endeavor, it seems that the corollary of this premise is research towards the end of ensuring these decisions conform to ethical norms."

"We need research into present and future AI that takes into account fundamental virtue ethics, deontological ethics, and consequentialist considerations."

"... Research in this area must in my opinion by definition be interdisciplinary, as no single discipline can hope to come to terms with the larger societal challenges of AI and autonomy. I will therefore try to compile a short (and decisively non-exhaustive) list of disciplines and forms of research that should come together here. (...) (1) Ethics. This might be the most obvious choice. Ethicists have the tools to reflect advances in technology development vis-a-vis larger trajectories of values and societal preferences, and can come up with points to consider."

"Research (is needed) on the evolving understanding of ethical implications of AI and autonomous machines."

FIGURE 4: EXAMPLE ASSERTION OUTPUT

2.5.5 Interim Results

The results in this section are the output of the analysis described above. It is included here for indication only, and no conclusions are drawn from these results as the consultation is not yet finished.

The results are presented in the form of the assertion statements (***bold italics***), together with how many participants expressed the statement, accompanied by original source quotations from the participants *(small italics)*.

[44] Qualitative research methods reliability of analysis is checked initially by checking agreement between two researchers ("coders") who attempt to identify categories and themes ("codes"). See, for example, Howitt, D. (2013) *Introduction to Qualitative Research Methods in Psychology*

[45] Any errors / national differences in spelling or grammar in the responses have been preserved to keep the responses authentic.

Co-funded by the Horizon 2020
Framework Programme of the European Union

Ethics

AI research and technical choices need to take into account ethical implications and norms

4 of 12 participants made this assertion in Round 1

Quotations from Round 1:

- *"As the premise of research in artificial intelligence and machine autonomy is increased of levels of automated decision making in all aspects of human endeavor, it seems that the corollary of this premise is research towards the end of ensuring these decisions conform to ethical norms."*
- *"We need research into present and future AI that takes into account fundamental virtue ethics, deontological ethics, and consequentialist considerations."*
- *"... Research in this area must in my opinion by definition be interdisciplinary, as no single discipline can hope to come to terms with the larger societal challenges of AI and autonomy. I will therefore try to compile a short (and decisively non-exhaustive) list of disciplines and forms of research that should come together here. (...) (1) Ethics. This might be the most obvious choice. Ethicists have the tools to reflect advances in technology development vis-a-vis larger trajectories of values and societal preferences, and can come up with points to consider."*
- *"Research (is needed) on the evolving understanding of ethical implications of AI and autonomous machines."*

Ethical choices and dilemmas faced by applications of AI need to be investigated, along with the factors that are relevant to them and the trade-offs between the factors in dilemma resolution

3 of 12 participants made this assertion in Round 1

Quotations from Round 1:

- *"What factors constrain the design of autonomous systems that can reliably behave in ways which are ethically acceptable?"*
- *"To facilitate autonomous ethical decision making, research should be undertaken regarding what ethically relevant features, and corollary duties, are present in the ethical dilemmas faced by these systems. Means of determining how choices in different contexts satisfy or violate these duties and principles that weigh these duties when they pull in different directions are also required."*
- *"(…) for many problems, there is significant diversity about the relevant values and their relative weights or importance. We thus need further research -- both empirical & normative -- about how to reconcile different value preferences & judgments."*

There is a deep inter-relationship between ethical and practical considerations in the design of autonomous machines

1 of 12 participants made this assertion in Round 1

Quotations from Round 1:

- *"Ethical and practical questions surrounding the design of autonomous machines are deeply intertwined."*

The concept of "autonomy" needs to be debated and agreed between philosophers, ethicists and engineers so as to come to a shared understanding about what types of machine qualify for ethical concern

1 of 12 participants made this assertion in Round 1

Co-funded by the Horizon 2020 Framework Programme of the European Union

Quotations from Round 1:

- *"... there is no clear understanding of the various senses of 'autonomy' used by philosophers and other ethicists (on the one hand) and engineers and robotics (on the other). As a result, it will be difficult to classify what machines (and which machine behaviors) fall into the category of 'autonomous machine'."*

Transparency

AI decisions need to be transparent, explained and justified

5 of 12 participants made this assertion in Round 1

Quotations from Round 1:

- *"Means to validate (ethical) features, duties, and principles need to be determined and the explanatory power of this data needs to be exploited to provide justification for the decisions made by such systems."*
- *"(What is needed is) transparency in algorithms (how to design for transparency in algorithms e.g. applied for decision making)"*
- *"What AI needs to do is to provide account of its decisions and reasoning patterns in order to support determine who is responsible. Thus explanation power and transparency are important areas of research."*
- *"Lack of transparency is inherent in several AI techniques, esp. in "big data" and in machine learning. It can generate problems of predictability, which means that systems must be tested and certified to a high degree if they are to be used in critical environments (where human lives are at risk)."*
- *"Research is needed to develop methods of governing autonomous machines which combine the flexibility of artificial neural networks and the trustworthiness of ethical algorithms drawn explicitly from human ethical theories. In particular, we need to develop techniques of machine learning that allow autonomous machines to justify or explain their decisions."*

AI decisions need to be understood by lay people, not just technical experts

3 of 12 participants made this assertion in Round 1

Quotations from Round 1:

- *"An increasingly important problem is the intelligibility of algorithms --both to their originators and to wider society."*
- *"How may non-technical professionals in policy, third sector or business spheres get to grips with the impacts of using algorithms in decision making. How may they counter some of their more discriminatory outputs and use them ethically? How may this be explained to users?"*
- *"Some degree of explainability or transparency appears to be quite valuable for beneficial AI, but we have only limited research into how to generate **psychologically useful** explanations. There is good cognitive science about what makes something a helpful explanation, but that work has not been translated into recommendations, practices, or theories for AI development & deployment."*

Transparency is needed for both data provenance and algorithmic decisions

1 of 12 participants made this assertion in Round 1

Quotations from Round 1:

- *"By transparency btw, I do not just mean making algorithms transparent, but more important is to make data transparent (use, collection, management, origin...). [Whilst it is important to make algorithms transparent, it is more important to make data transparent, e.g. provenance (use, collection, management, origin)] "*

Transparency requires that the autonomous machines meet at least two criteria: a track record of reliability and comprehensibility of previous behaviour

1 of 12 participants made this assertion in Round 1

Quotations from Round 1:

- *"To avoid distrust and to allow proper consent, responsible autonomous machines must be appropriately transparent. Transparency requires that the autonomous machines meet at least two criteria: track record of reliability and comprehensibility of previous behaviour. A track record of reliability requires an autonomous machine to have been tested in a range of stimulations and found to make acceptable ethical decisions. But a track record of reliability is not enough, transparency, and genuine consent requires that we have some grasp on how the system reaches its decisions."*

Regulation & Control

An EU AI and Robotics Social Ethics committee should be formed as a priority over AI technical research

1 of 12 participants made this assertion in Round 1

Quotations from Round 1:

- *"The issues raised by "Responsible Autonomous Machines and Beneficial AI" are foremostlty ideological and political in a deep way: who benefits, who lose jobs, who oversees and regulates that. Throwing research money at them without addressing those aspects is actually a way of deliberately masking the existence of such deep issues, and of reducing them to technicalities and technical progress not to be confused with social progress however. Just like there are EU Bio-Ethics committees, an EU AI and Robotics Social Ethics committee should be formed beforehand to publicly frame and express the deeper concerns"*

Interdisciplinary research is needed to determine how law can ensure responsible behaviour

2 of 12 participants made this assertion in Round 1

Quotations from Round 1:

- *"Only when all relevant perspectives are on the table can legal regulation accommodate the multi-dimensional requirements that technical systems prescribe."*
- *"Interdisciplinary research is very much needed in this field to (...) determine how public and/or private law can ensure responsible behaviour"*

Mechanisms that monitor and can constrain AI systems' behaviour are necessary, including stop or human override controls

3 of 12 participants made this assertion in Round 1

Quotations from Round 1:

- *"Another area of research wrt transparency is to develop monitoring and control mechanisms that guarantee which are the limits of any algorithm so that one knows what it can, and most importantly, what it cannot do."*
- *"Clearly machines learning to achieve objectives can pursue them in counter-intuitive ways, and it is desirable that counter-intuitive actions be controllable --if only by shutting a system down. So research on stop mechanisms is important."*
- *"What control mechanisms will enable autonomous machines to behave in ethically acceptable ways in human environments?"*

Research into whether and how AI systems' decisions or actions can be rolled back is needed

1 of 12 participants made this assertion in Round 1

Quotations from Round 1:

- *"… So is research on reversibility --on the model of the restore to earlier state software on PC and laptop operating systems."*

Certification of reliably safe AI is needed, including definitions of criteria for safety

3 of 12 participants made this assertion in Round 1

Quotations from Round 1:

- *"In what ways can it be adequately demonstrated that an autonomous machine is reliable enough to be allowed to operate in an uncontrolled human environment?"*
- *"What standard of reliability should we impose on autonomous machines?"*
- *"To what extent can certification schemes be applied to algorithms, so that business organizations using them can demonstrate to their stakeholders that they are not doing something discriminatory or unethical, whilst retaining commercial secrecy?"*
- *"Lack of transparency is inherent in several AI techniques, esp. in "big data" and in machine learning. It can generate problems of predictability, which means that systems must be tested and certified to a high degree if they are to be used in critical environments (where human lives are at risk)."*

Research is needed to analyse the factors that lead to liability

1 of 12 participants made this assertion in Round 1

Quotations from Round 1:

- *"Analysing the factors that lead to liability (...) are crucial steps to be followed within each relevant sector."*

Penalties need to be identified for liable parties associated with autonomous machines that transgress

1 of 12 participants made this assertion in Round 1

Quotations from Round 1:

- *"Identifying (...) the remedies attached to liability rules are crucial steps to be followed within each relevant sector."*

Co-funded by the Horizon 2020 Framework Programme of the European Union

Social Impact

Research into AI's impact on human workers is needed, including employment and deskilling of humans replaced by machines, as well as psychological consequences

2 of 12 participants made this assertion in Round 1

Quotations from Round 1:

- *"The issues raised by "Responsible Autonomous Machines and Beneficial AI" are foremostlty ideological and political in a deep way: who benefits, who lose jobs, who oversees and regulates that."*
- *"The implications for employment of autonomous cars and delivery drones need extensive research."*
- *"The health and psychological implications of mass displacement of people by machines"*
- *"(Research is needed on) the deskilling of highly trained professionals who are overdependent on e.g. diagnostic machines or robot surgery need research."*

Research into the threats that future AI may pose to humankind is required, including where AI and human goals differ and where AI can undermine human values

2 of 12 participants made this assertion in Round 1

Quotations from Round 1:

- *"We need to know whether we should expect independent agents based on AI, i.e. systems that individually try to achieve goals and pursue these, even if they might be in contradiction to human goals - perhaps even systems that can be said to be responsible for their actions."*
- *"There is agreement that autonomous and highly intelligent AI could, in principle, constitute an existential threat to humankind, but this is usually seen as a theoretical or very long-term possibility only. Given that the consequences could be of great importance, even a small probability of them coming about is sufficient to warrant research into these questions as well."*
- *"More autonomous robots may lead to less human control, and in the long run they may lead to a situation that is not beneficial to humankind. Some of the economic results of robot use in automation are also seen as ethical problems, e.g. changes in labour conditions, loss of jobs or a more uneven distribution of wealth."*
- *"The exponential growth of computing power and some advances in AI algorithms will continue to lead to rapid development in AI and robotics for at least a few decades. This development has the potential of undermining human values, esp. moral responsibility ("the robot did it!"), compassion, and human dignity."*
- *"Future and present AI and robotics will have negative consequences on the well-being of humans (and other sentient beings) in many ways. This includes safety at the workplace, de-humanisation of certain environments (such as health-care), and easier killing of humans in war. Here the question is: Are the benefits of robots worth the risks? And: Are the risks distributed fairly?"*
- *"Research is needed on how mass failure of industrial machines and autonomous vehicles could be survived and reversed. The fact that mass failure could be catastrophic implies the need for research on safe networking of the relevant objects. Decentralization of networks and reduced interoperability might be desirable on security grounds."*

Public attitudes towards AI need to be understood, especially concerning public trust of AI

1 of 12 participants made this assertion in Round 1

Quotations from Round 1:

- *"Moral philosophers should be involved in interdisciplinary research to help acquire a better understanding of public attitudes towards the behaviour of autonomous machines. (...) Moral philosophers could be helpful in helping to develop the right kinds of cases for empirical researchers to present to the public to give us a better understanding of the types of factors that they see as morally relevant."*
- *"As well as compelling ethical reasons to ensure that autonomous machines make good and justifiable decisions, there are prudential reasons to do so. Distrust of autonomous machines amongst the public is likely to hold back their implementation."*

Research is needed into how AI integrates into networks of humans and machines, as well [as] how machines interact with other machines

2 of 12 participants made this assertion in Round 1

Quotations from Round 1:

- *"Research on the interplay between AI/autonomous machines and humans, in human-machine networks. E.g. the constructive interplay between humans and automnomous machines (the integration of autonomous machines in teams), and the constructive interplay between networked autonomous machines."*
- *"We certainly need to make sure that autonomous machines work to the benefit of humanity. Hence sociological and political research into the effects of adopting ever more autonomous systems is pressing."*

Research is needed into how users of AI can identify and guard against discriminatory effects of AI

1 of 12 participants made this assertion in Round 1

Quotations from Round 1:

- *"How may partners working with disadvantaged groups ensure that they are empowered so that they do not suffer cumulative disadvantage because of the discriminatory effects of algorithms (e.g. when used in criminal justice settings)."*

We need to understand the economic motivations behind technical systems that operate within society

1 of 12 participants made this assertion in Round 1

Quotations from Round 1:

- *"Technological developments are usually outcomes from research and development that serves commercial interests. For an understanding of the impacts of advanced technical systems in society, it is therefore important to have knowledge about their origins and purposes, and the economic rationales that underpin these systems."*

AI systems cannot set their own goals or motivations

1 of 12 participants made this assertion in Round 1

Quotations from Round 1:

Co-funded by the Horizon 2020
Framework Programme of the European Union

- *"AI systems (current and most likely for the coming future) are only autonomous wrt to their plans. They cannot set their own goals, let alone their motives."*

Research is needed into how AI can be tested against societal values such as self-determination, autonomy, freedom, trust and privacy

1 of 12 participants made this assertion in Round 1

Quotations from Round 1:

- *"To what extent can privacy be conceptualised as a social value i.e. as something that can benefit society and other values such as self-determination, autonomy, freedom, trust etc. How can algorithms be challenged on these grounds?"*

Research is needed to explore the differences between decisions made by autonomous machines and humans in the same situation

1 of 12 participants made this assertion in Round 1

Quotations from Round 1:

- *"Research is also required to explore the differences between decisions made by autonomous machines and analogous situations involving human agents. Some ethical considerations that apply to the actions of human agents are not likely to apply to autonomous machines. For example, when a human driver is in a collision situation, they are likely to face serious challenges in making the right decision under pressure; we may also not be able to expect the drivers to behave in certain ways because it is either psychologically or physically difficult. Autonomous machines will not face the same excuses related to difficulty (for) human drivers."*

"Beneficial AI" is a poor term because it is not clear who benefits, and it does not account for any harm that may occur to others.

1 of 12 participants made this assertion in Round 1

Quotations from Round 1:

- *"All AI is beneficial for something. At the very least for its researchers to publish some paper. Most likely to the companies using it. So what are you talking about when you claim beneficial AI? the whole of mankind? This would mean everybody, the 'good' and the 'bad'. Do you want AI to benefit e.g. ISIS or other terrorists? And if is not really everybody, then who gets to decide who benefits and who doesn't? And why should we accept such decisions? Isn't it what corporations are doing now already? They decide who benefits (mostly their shareholders) Why would you do better?"*

Design

Ethical principles need to be embedded into AI development

2 of 12 participants made this assertion in Round 1

Quotations from Round 1:

- *"Development of AI & autonomous technologies almost always requires specification & incorporation of values from the outset. People often focus on values & ethics as things that are added at the end of development (e.g., through a module that can "veto" proposed actions). But development of an AI system almost always requires specifying a loss or value function to be optimized, and that means that we need to specify (at least partially) the relevant values -- that is, what should "matter" for the AI system."*

- *"Where are values and interests embedded into algorithm production processes?"*

AI engineers need to be aware of potential biases and prejudices in selection of training data

1 of 12 participants made this assertion in Round 1

Quotations from Round 1:

- *"Starting from the assumption that AI will in the future continue to hinge on training data, engineers must be conscious about certain social assumptions (e.g. stereotypes, clichés, prejudices, etc.) that go into technical systems, and need to strongly reflect on ways how to possibly avoid structural discrimination."*

Inclusive, interdisciplinary teams are needed to develop AI

2 of 12 participants made this assertion in Round 1

Quotations from Round 1:

- *"Finally, one needs to look at the processes of AI development. Ensure inclusive, diverse teams (…)"*
- *"Research in this area must in my opinion by definition be interdisciplinary, as no single discipline can hope to come to terms with the larger societal challenges of AI and autonomy. I will therefore try to compile a short (and decisively non-exhaustive) list of disciplines and forms of research that should come together here. (...)*
    - *(1) Ethics (...)
(2) Engineering and computer sciences (...)
(3) Social sciences (...)
(4) Economics (...)
(5) Law (...)"*
- *"Where ethics operates on an abstract level, and engineers and computer scientists are concerned with the development of advanced technical systems in the lab, social sciences offer the much-needed perspective on what happens with technologies "in the wild"."*

Formal definitions of all concepts are necessary to avoid ambiguity and unnecessary concern

1 of 12 participants made this assertion in Round 1

Quotations from Round 1:

- *"First important issue is to provide formal definitions of all concepts. What is meant by 'beneficial', 'responsibility' and 'autonomy'. All these concepts have been used loosely and inconsistently in scientific and newspaper pieces alike, contributing to the feelings of unease wrt AI."*

Design of AI needs a human-centred approach

1 of 12 participants made this assertion in Round 1

Quotations from Round 1:

- *"(What is needed is) human-centred design of AI and autonomous machines (how to involve users / stakeholders in design process?)"*

Responsibility

People, not AI systems, bear responsibility and AI developers are responsible for the tools they develop

2 of 12 participants made this assertion in Round 1

Quotations from Round 1:

- *"Engineers and computer scientists bear major responsibilities for the tools that they develop, and therefore need to make conscious choices when designing technical systems."*
- *"We strongly place responsibility with people. AI systems are artefacts, build for some reason, by someone. AI systems are tools."*

The concept of "AI responsibility" needs to be researched by integrated, multidisciplinary teams so as to arrive at a hybrid understanding of the key issues concerning responsibility and where it can be attributed when AI participates in human-machine networks

5 of 12 participants made this assertion in Round 1

Quotations from Round 1:

- *"As many algorithms are open source, produced internationally by disparate and unconnected groups of computer scientists, who is responsible for the ethicality of inputs in algorithm training and testing?"*
- *"There is considerable complexity in the very notion of responsibility as applied to technology. Is it a causal account? Does it allow for distributed responsibility over individuals and groups of humans and technologies? Philosophers need to understand that these two notions--autonomy and responsibility--are not the exclusive domain of philosophy. Lay people, and engineers and scientists, have their own conceptions which are "operationalized" in many different contexts. I propose that some necessary research on developing hybrid and integrated conceptions of (autonomy and responsibility) will go a long way (but not all the way) towards being able to address the aforementioned issues."*
- *"How responsibility -- moral, causal, social, and other -- is & should be allocated among members of a team, particularly when those members have different roles or functions. I frame this question without mentioning AI or autonomous technologies, but obviously the key "use case" is when we have a human-machine team (perhaps multiple humans, perhaps multiple machines)."*
- *"What mechanisms of accountability can be embedded into such control systems?"*
- *"Interdisciplinary research is very much needed in this field to conceptualise the notion of responsibility"*
- *"Identifying the bearer of the duty (...) are crucial steps to be followed within each relevant sector."*

Research is needed to determine how/when responsibility should translate into liability

1 of 12 participants made this assertion in Round 1

Quotations from Round 1:

- *"Interdisciplinary research is very much needed in this field to (...) determine (...) how/when responsibility should translate into liability."*

2.6 ECHO CHAMBERS & FAKE NEWS

The justification for the theme of Echo Chambers & Fake News is given in D2.1 and the NGI Emerging Research Challenges White Paper[46]. The white paper states:

> *Many sources agreed that there is a risk that the Internet becomes an "echo chamber", where profiling of citizens; and citizens' preferences and social groups limit the information they can see to sympathetic views, reinforcing the citizens' entrenched views.*
>
> *Multidisciplinary research is needed in order to answer questions relating to the promotion of diversity and truth in the Internet. Many of these questions relate to the causes of limited or biased information and how the information can be made less biased or more complete. Examples of causes include unbalanced search results from Internet search providers that tune the results to users' previous searches or preferences; restrictions on Internet search results through interventions by authoritarian governments; the current high-profile of "fake news" (is the news really fake or is someone merely accusing it of being fake?); and social groups that pursue a particular agenda by reinforcing certain arguments, ignoring other opinions.*
>
> *These questions raise other questions of jurisdiction, state control and liberty, and a question overarching them all is: what levels of intervention are acceptable before liberty is compromised?*

2.6.1 Knowledge Requirements & Expert Selection

Knowledge requirements were determined in a similar manner to the Responsible Autonomous Machines consultation (reported here in Section 2.5.1) – the initial set of themes from D2.1 was used as starting points for an exploratory literature survey to determine related themes. The initial set of themes was:

- Internet echo chambers
- Fake news & information verification
- Bias in information
- Closed communities in the Internet

A literature survey was conducted to determine additional related themes, and these are listed as follows:

- Homophily
- Filter bubbles
- Search engine bias & manipulation of results
- Search engine manipulation
- Information credibility
- Social networking sites' impact on truth
- Journalism & media

[46] Taylor, S., Boniface M. Next Generation Internet: The Emerging Research Challenges - Key Issues Arising from Multiple Consultations Concerning the Next Generation of the Internet. https://ec.europa.eu/futurium/en/next-generation-internet/next-generation-internet-emerging-research-challenges-key-issues-arising

- Propaganda in the Internet
- Internet censorship
- User profiling by search providers
- Internet information provenance

These themes were used as search terms in different investigations, in a similar fashion to that described in Section 2.5.1. The literature survey also served to provide source material for a background briefing note for potential consultees, which is presented in the next section.

2.6.2 Status

The current status is that the literature survey and the expert selection for the Echo Chambers & Fake News consultation is a work in progress, and the next section contains a first draft of the background briefing paper. The consultation is planned for launch in mid-January 2018.

2.6.3 Background

This section provides a first draft version of the background to the Echo Chambers consultation. It may be updated as necessary for the actual launch of the consultation in Q1 2018.

ECHO CHAMBERS & FAKE NEWS – CONSULTATION WITH EXPERTS: Background & Gateway Questions

The Internet provides citizens with easy access to vast amounts of information at the touch of a button, but the information is not necessarily verified and may present a distorted view of real events or facts. There is a risk that the Internet becomes an "echo chamber", where profiling of citizens and citizens' preferences and social groups limit the information they can see to sympathetic views, reinforcing the citizens' entrenched views.

Multidisciplinary research and innovation are needed in order to answer questions relating to the promotion of diversity and truth in the Internet. Many of these questions relate to the causes of limited or biased information and how the information can be made less biased or more complete. Examples of causes include unbalanced search results from Internet search providers that tune the results to users' previous searches or preferences; restrictions on Internet search results through interventions by authoritarian governments; the current high-profile of "fake news" (is the news really fake or is someone merely accusing it of being fake?); and social groups that pursue a particular agenda by reinforcing certain arguments, ignoring other opinions.

The phrase "filter bubble" was coined by Eli Pariser[47] and refers to the isolation of citizens in "bubbles" of information filtered to suit their opinions. A 2016 consultation by the Ditchley Foundation stated that *"there is a risk that the Internet becomes an echo chamber for our own prejudices and preconceptions, rather than a source of objective facts and challenge. We are already seeing this in the rapid spread of false news"*[48]. Influential figures such as Bill Gates have identified the dangers of closed communities that reinforce entrenched opinions: *"[Technology such as social media] lets you go off with like-minded people, so you're not mixing and sharing and understanding other points of view ... It's super important. It's turned out to be more of a problem than I, or many others, would have expected"*[49].

Internet content can be filtered and censored, often without the knowledge of the consuming citizens. The Ditchley Foundation consultation states that *"increasingly consumers are being presented with a selected slice of the Internet, controlled, filtered and sanitised"*[48]. Internet

[47] Eli Pariser. The Filter Bubble: What the Internet Is Hiding from You. 2011.

[48] Will we still have a single global Internet in 2025? - The Ditchley Foundation http://www.ditchley.co.uk/conferences/past-programme/2010-2019/2016/global-Internet 2016

[49] Filter bubbles are a serious problem with news, says Bill Gates. https://qz.com/913114/bill-gates-says-filter-bubbles-are-a-serious-problem-with-news/

Co-funded by the Horizon 2020 Framework Programme of the European Union

search result bias, where different users get different search results for the same query based on the search provider's profiling of the user and advertisement targeting, is not new: in 2005, Goldman stated: *"Due to search engines' automated operations, people often assume that search engines display search results neutrally and without bias. However, this perception is mistaken. Like any other media company, search engines affirmatively control their users' experiences, which has the consequence of skewing search results (a phenomenon called 'search engine bias')"*[50]. Carson appeals to Google to provide a switch *"that will allow users to manually toggle between results returned through Google's new personalization algorithms and results returned through Google's original PageRank algorithms"*[51] so as to show the effects of the personalization algorithms and enable users to avoid filter bubbles.

The issue of misinformation and fake news is clearly becoming highly important. Whilst propaganda has been around for centuries, the ease with which false & biased information can be spread, coupled with the current perceived magnitude of its impact, means that research into addressing the issues of misinformation and fake news is becoming pressing. Tim Berners-Lee states that the web needs *"saving"*, and major issues to be addressed are that *"It's too easy for misinformation to spread on the web"* and *"political advertising online needs transparency and understanding"*[52,53]. Whilst not claiming that fake news affected the outcome, a recent study into fake news and its impact on the 2016 US election by Allcott & Gentzkow indicated that number of false news stories shared on Facebook favouring Trump was 3.75 times greater than those favouring Clinton: *"of the known false news stories that appeared in the three months before the election, those favoring Trump were shared a total of 30 million times on Facebook, while those favoring Clinton were shared 8 million times"*[54].

Information credibility is a critical factor. Citing previous work, Johnson and Kaye[55] consider the question of Internet information credibility from the perspective of trust placed in political information from social media networks, and found that *"politically interested Internet users in general judged SNS quite low in credibility, 7.4 on a 4–20 point index"*. As with trust, credibility is a judgement made by the recipient of information possibly based on many factors. Metzger[56] states that "*a long history of research finds that credibility is a multifaceted concept with two primary dimensions: expertise and trustworthiness"*. Bias and propaganda have been present in traditional media (e.g. newspapers and TV), especially if the broadcaster is not independent from political or other driving forces (e.g. a state-controlled press). When a free press exists, broadcasters trade on their reputation and consumers place their trust in the broadcasters to provide accurate reports of world events. Much traditional media is subject to basic journalistic practices such as fact-checking because publishers have a vested interest in protecting their reputation and avoid libel suits by checking the information they broadcast is factually correct. Even then, citizens can be subject to echo chambers through the broadcast media they choose to receive news through, e.g. left-leaning citizens are more likely to read leftist newspapers. Whilst these channels still exist today, there is an additional deluge of information - nowadays it is easy for anyone to publish unverified information or propaganda in the multitude of channels and locations available today in the Internet, and any bias or slant is very often not explicit. Metzger also discusses user motivation whether to determine credibility or accept information at face value, and makes the distinction between different users making different judgements whether to assess the credibility of information in the Internet carefully. The issue of motivation

[50] Goldman, E., 2005. Search engine bias and the demise of search engine utopianism. Yale JL & Tech., 8, p.188.

[51] Carson, A.B., 2015. Public Discourse in the Age of Personalization: Psychological Explanations and Political Implications of Search Engine Bias and the Filter Bubble. Journal of Science Policy & Governance, 7(1).

[52] Tim Berners-Lee, I invented the web. Here are three things we need to change to save it. The Guardian, 12 March 2017. https://www.theguardian.com/technology/2017/mar/11/tim-berners-lee-web-inventor-save-internet

[53] Berners-Lee also mentions that citizens have lost control of their personal data. This is important, but is outside the scope of this consultation as it is already well addressed by existing EC work programmes.

[54] Allcott, H. and Gentzkow, M., 2017. Social media and fake news in the 2016 election (No. w23089). National Bureau of Economic Research. © 2017 by Hunt Allcott and Matthew Gentzkow.

[55] Johnson, T.J. and Kaye, B.K., 2014. Credibility of social network sites for political information among politically interested Internet users. Journal of Computer-Mediated Communication, 19(4), pp.957-974.

[56] Metzger, M.J., 2007. Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. Journal of the Association for Information Science and Technology, 58(13), pp.2078-2091.

to evaluate credibility touches on the previous subtopic of filter bubbles and entrenched opinions: whether the information recipient is motivated to evaluate the information may also depend on their own opinions and biases. For example, are citizens likely to be critical of the information they find in the Internet if it agrees with their world view?

These questions raise other questions of jurisdiction, state control and liberty, and a question overarching them all is: what kinds and levels of intervention are acceptable before liberty is compromised?

Following are starting point questions, aimed at kicking off the consultation process. They are by nature open-ended, so as to give respondents maximum freedom in their input.

- ***What research is needed to address the issues that Echo Chambers raise?***
- ***Why is the recommended research important?***
- ***What types of resources (both human e.g. skills or expertise; and tools, e.g. surveys or computational resources) are needed for the research?***

3 NGI INNOVATION PATHWAYS

3.1 INTRODUCTION

This part of the Deliverable D2.2 sets the background for innovation pathways which might be used for NGI innovators to work towards the successful accomplishment of their goals. The outcomes of the consultation activities described previously are beginning to inform the development of a domain model for innovation for the NGI, allowing us to explore different innovation pathways. These pathways will then need to be evaluated in the context of work in other work packages to allow us to revise our initial models as needed and move forward to support initial testing in WP3. To begin with, we offer a number of relevant definitions before moving on to describe the domain model we propose, how this relates to and has been informed by consultation work to date, and how we currently plan to evaluate this work further.

3.2 DEFINITION

Many different definitions of *Innovation* have been offered[57]. In typically gnomic fashion, the Oxford English Dictionary, for example, suggests:

Innovation	[mass noun]	The action or process of innovating. ‘*innovation is crucial to the continuing success of any organization*
	[count noun]	A new method, idea, product, etc. ‘*technological innovations designed to save energy*’[58]

Albeit distinguished by the type of noun, such definitions conflate *process* and *product*. Similarly, the online Business Dictionary describes innovation in the following terms:

> “The *process* of translating an idea or invention into a good or service that creates value or for which customers will pay.
>
> To be called an innovation, an *idea* must be replicable at an economical cost and must satisfy a specific need. Innovation involves deliberate *application* of information, imagination and initiative in deriving greater or different values from resources, and includes all processes by which new ideas are generated and converted into useful products. In business, innovation often results when ideas are applied by the company in order to further satisfy the needs and expectations of the customers”[59]. (*Our emphasis*)

As well as mixing together *process* and *outcome,* this definition introduces two further notions. First, the innovation should “create value” for someone; and secondly, it may result from applying something already known in a novel way. The focus here is on outcome as much as novelty. Professor of Innovation and Entrepreneurship at the University of Exeter, John Bessant, takes this a little further. His definition runs:

> “Innovation means *creating value from ideas*. While a lot of interest is in commercial value, a lot can be done with *social value*. For the Red Cross, creating social value is a case of life and death, and while it’s not creating lots of

[57] See, for example, https://www.ideatovalue.com/inno/nickskillicorn/2016/03/innovation-15-experts-share-innovation-definition/

[58] https://en.oxforddictionaries.com/definition/innovation ; for completeness, see also: https://dictionary.cambridge.org/dictionary/english/innovation

[59] http://www.businessdictionary.com/definition/innovation.html

> money, it's creating real value from ideas, such as simple low-cost hygiene products to avoid sanitation-linked infection.
>
> *There should be no limits as to where innovation comes from*. It can come from our own teams, what competitors are doing, and the market. Today, *it's all about what users want and need*, so it's up to businesses to make sure that they have *a good set of antennae* to pick up on these trends"[60]. (*Our emphasis*)

Once again stressing *value*, which is expanded beyond purely commercial terms to include societal benefit, this definition also introduces the concept that innovation may come from anywhere and is motivated by user needs and wants. There must therefore be some openness to look for input and understanding from a whole range of disciplines and domains ("a good set of antennae").

The concept of *innovation* therefore may include an outcome on its own, with societal as well as commercial benefit, may come from any area, and may involve a process to take an idea or motivation and turn it into a suitable result. For the purposes of this discussion, we will try to distinguish outcome and process and define innovation as follows:

Innovation process	The steps taken to convert an idea, need or want into a novel solution
Innovation	The outcome of the *Innovation process*

With these working definitions, we will develop an understanding of the innovation domain in the following sections which attempts to incorporate the motivation and sources for innovation, the environment in which the innovation process takes place, and where and how the innovation outcome is delivered. In so doing, we seek to enable ***Innovation Pathways***, which we define as the set of steps needed the *innovation process* a success as well as the experts which need to be involved.

3.3 RATIONALE

Deliverable D2.1 identified a set of nine challenges which need to be addressed as the Internet landscape changes in the future. It is not enough, however, simply to identify what those challenges are. Instead we need to identify what is available and what is missing to be able to manage those challenges and in so doing describe how innovation might be enabled. With this in mind, this deliverable is intended to develop an initial understanding of how innovation pathways may be elaborated within the innovation space. As a first step, we need to understand the individual constructs involved in an innovation pathway and the expertise required to enable that pathway.

This section of the deliverable will therefore provide the background which will take us forward to identify what gaps exist in the innovation space, and to develop a process to configure the environment required to address one or two of the challenges identified.

3.4 TARGET AUDIENCE

The Innovation Pathway section of the deliverable is therefore aimed at:

[60] http://www.telegraph.co.uk/connect/better-business/innovation/what-is-innovation-and-how-can-businesses-foster-it/

- *Other project partners*: so that we can share our findings across work packages in support of the complementary activities in those work packages;
- *Other stakeholders in the project*: as part of the results and outcomes to be made available as part of dissemination activities to parties such as the EC, and other projects interested in the NGI; and therefore
- *Other stakeholders in the NGI*: any other group, project or individual who has an interest in where we believe the NGI is progressing, and how we may support development and technology advancement.

The following sections begin with a description of the innovation space as we see it, followed by the next steps after what is presented here which will position this deliverable in regard to other work packages.

3.5 DOMAIN MODEL

This section introduces the main constructs which we define within our definition of *Innovation Process*. Beginning with the high-level domain model, further subsections describe individual constructs and how they relate to the domain.

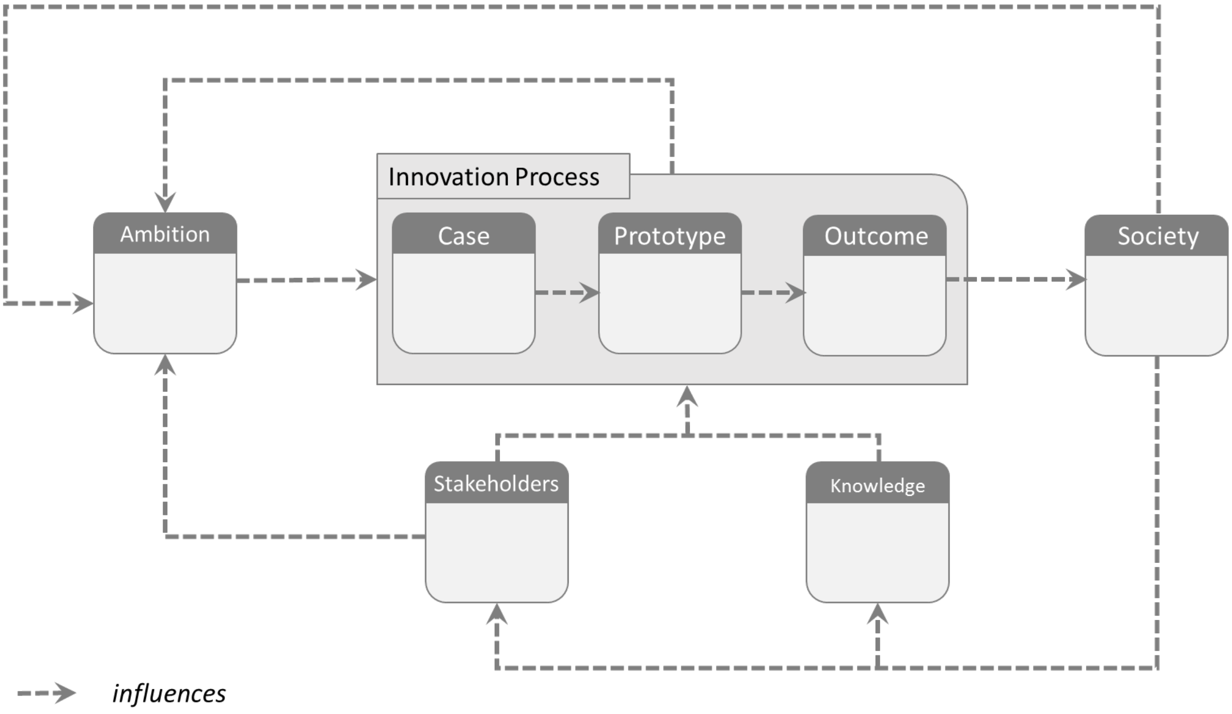


FIGURE 5: HIGH LEVEL DOMAIN MODEL FOR INNOVATION

The overall *Innovation Process* is contextualised within the broader ecosystem in Figure 5. Individual constructs are described in more detail in the following sections. There are four main components:

1. The *Ambition* is the overall vision, what needs to be achieved; and defines the rationale and motivation behind a given innovation. Note that the *Ambition* may be influenced both by *Society* as a whole, what it expects and what it will not accept, as well as *Stakeholder* perceptions and expectations. This feeds into:
2. The *Innovation Process* itself where an idea is evaluated and implemented or elaborated to produce a recognisable *Outcome*. The process is described in more detail below (Section 3.5.1). As well as responding to an *Ambition,* the *Innovation Process* is informed and constrained by two constructs:

3. On the one hand, there are *Stakeholders* who have an interest in whatever the *Outcome* of the process, but also in how it is achieved. *Stakeholders* are described in more detail in Section 3.5.6 below. On the other hand, *Knowledge* in broad and general terms (see Section 3.5.4) will constrain what can be done and influence the choices on how it can be done.
4. Finally, *Society* is the main beneficiary of the *Innovation Outcome*, but may also constrain it (via *Knowledge* and *Stakeholders*) or seed innovation (via *Ambition*).

Note that the convention in the figures in this document is to use a hashed arrow in association with the concept of "influence" or "leads on to"; and arrow leading to multiple constructs (such as the one from *Stakeholders / Knowledge* to the *Innovation Process*) may influence any individual or all of the constructs within the grouping. Similarly, a single arrow from a grouping (e.g., from *Innovation Process* to *Ambition*) means that any of the individual constructs within the grouping may affect the construct at the end of the arrow.

3.5.1 The Innovation Process

As mentioned above, in response to a suitable *Ambition* (Section 3.5.3) an *Innovation Process* may be initiated with the hope that it will lead to an appropriate change (*Outcome*) with some benefit to or at least effect on *Society*.

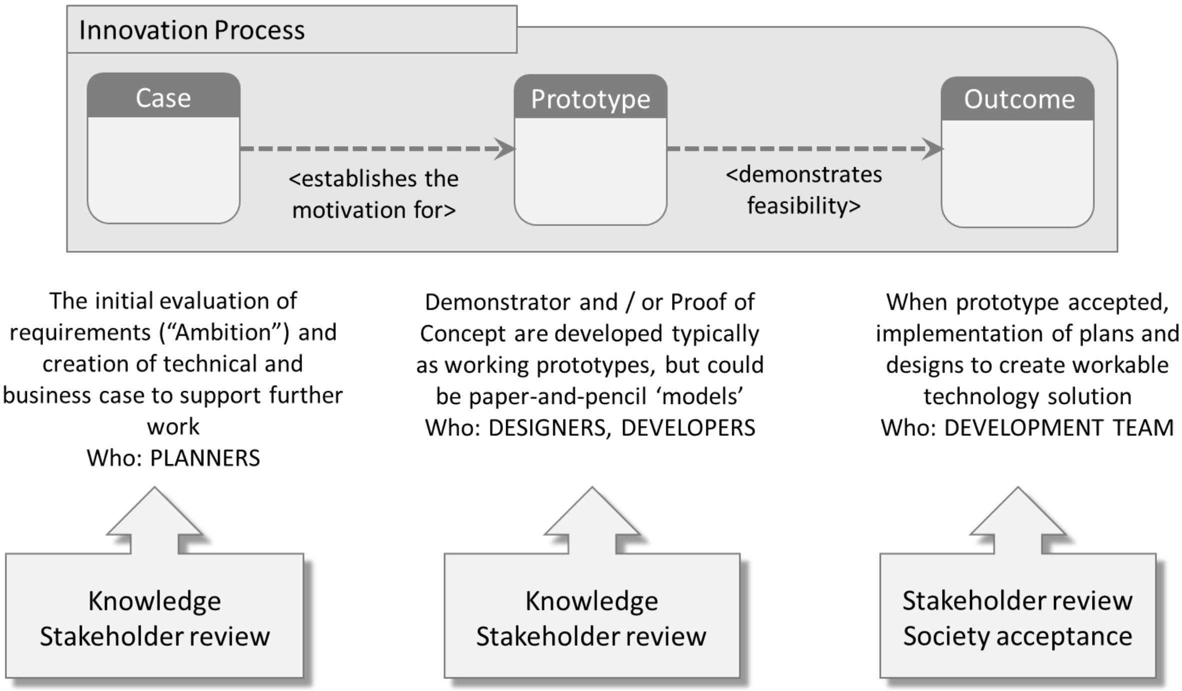

Innovation Process
Case
<establishes the motivation for>
Prototype
<demonstrates feasibility>
Outcome
The initial evaluation of requirements ("Ambition") and creation of technical and business case to support further work
Who: PLANNERS
Demonstrator and / or Proof of Concept are developed typically as working prototypes, but could be paper-and-pencil 'models'
Who: DESIGNERS, DEVELOPERS
When prototype accepted, implementation of plans and designs to create workable technology solution
Who: DEVELOPMENT TEAM
Knowledge
Stakeholder review
Knowledge
Stakeholder review
Stakeholder review
Society acceptance

FIGURE 6: OVERVIEW OF THE INNOVATION PROCESS

The *Innovation Process* itself (Figure 6) depends on a suitable business and technical justification (or *Case*, below) which justifies the effort involved in taking an idea further. Existing *Knowledge* (Section 3.5.4) may influence the development of the *Case*; and *Stakeholders* (Section 3.5.6) would review the *Case* and possibly how the *Case* was developed. Within the *Innovation Process*, there are a number of actors involved exclusively with the process itself, as opposed to the *Stakeholders,* who are external to it. These include:

- *Designers:* those actors responsible for generating a specification from which the *Developers* would work, and in response to what they have been told as the overall aims and requirements;
- *Developers:* who implement the design(s) from the *Designers*;

- *Development teams:* groups of individuals who support the handover of the *Innovation Outcome* to *Society*; they would include *Developers* at least, but may also rely on *Tester* and *Support* and/or *Service* experts; and
- *Planners*: who build up the non-technical overview of what will need to be implemented to support the *Outcome*.

At this stage, it is mainly *Planners* who are involved in the evaluation of the *Ambition* and its development into a suitable *Case* to justify the effort and work to be undertaken.

A successful *Case* would motivate continued effort within the *Innovation Process*. The idea or plans would be developed into a *Prototype*, or proof of concept, to demonstrate that it is possible to implement the idea. Once more, prior *Knowledge* would influence how the *Prototype* is developed, even if that involves a complete paradigm shift via the rejection of what has been done previously. Similarly, *Stakeholders* would monitor progress and what is being done. *Designers* and *Developers* would work together at this stage to demonstrate the feasibility and implementation of the ideas coming from the original *Ambition,* mediated by the *Case.*

Finally, the *Innovation Process* completes as an *Outcome* is achieved: an idea, a product, or a process. This would continue to be monitored and overseen by the relevant *Stakeholders*; and would need to be evaluated to determine acceptance by *Society*. Now a *Development Team*, which would include supporting roles such as test, promotion, and support for instance, is responsible for making the handover to *Society* happen.

3.5.2 Identification of key skills

In developing innovation pathways, it is essential to be able to identify what skills might be needed to support any given construct within the domain model here. For the purposes of this discussion, we will assume that the innovation outcome relates to technology. By default, therefore, most of the constructs in the model would require input and support from ICT planners, designers, developers and engineers. The constructs *Society* and *Stakeholders* would involve all relevant parties, including individual citizens.

For some of the other constructs, though, specific expertise is required. In the following sections, therefore, we have identified appropriate skills that should be engaged to support a given task or activity. These include:

- *Ethics*: expertise in understanding ethical implications associated with a given event, process or device (the innovation); and occasionally
  - *Philosophy*: the epistemological framework upon which a given innovation may rely;
- *Legal*: expertise in current and possible regulation which might relate to the innovation;
- *Sociology*: expertise in all aspects of social interaction and societal structures;
- *Psychology*: expertise in the understanding of individual attitudes and behaviours, including online (cyberpsychology) by comparison to real-world activities;
- *Economics*: understanding of the financial and other benefit structures which are relevant to a given innovation;
- *(Human) Geography*: expertise in the physical environment and how this affects and is affected by the innovation;
- *(Web) Social Science*: expertise specifically in online activity enablement and data management and sharing.

Relevant experts from these fields would have to collaborate to ensure the successful execution of a given pathway.

Co-funded by the Horizon 2020
Framework Programme of the European Union

3.5.3 The constructs associated with AMBITION

As previously stated, the *Innovation Process* is initiated in response to an *Ambition*, or the sum of user needs, existing technology and context requirements, and previous *Knowledge*. Figure 7 summarises the constructs associated with *Ambition*.

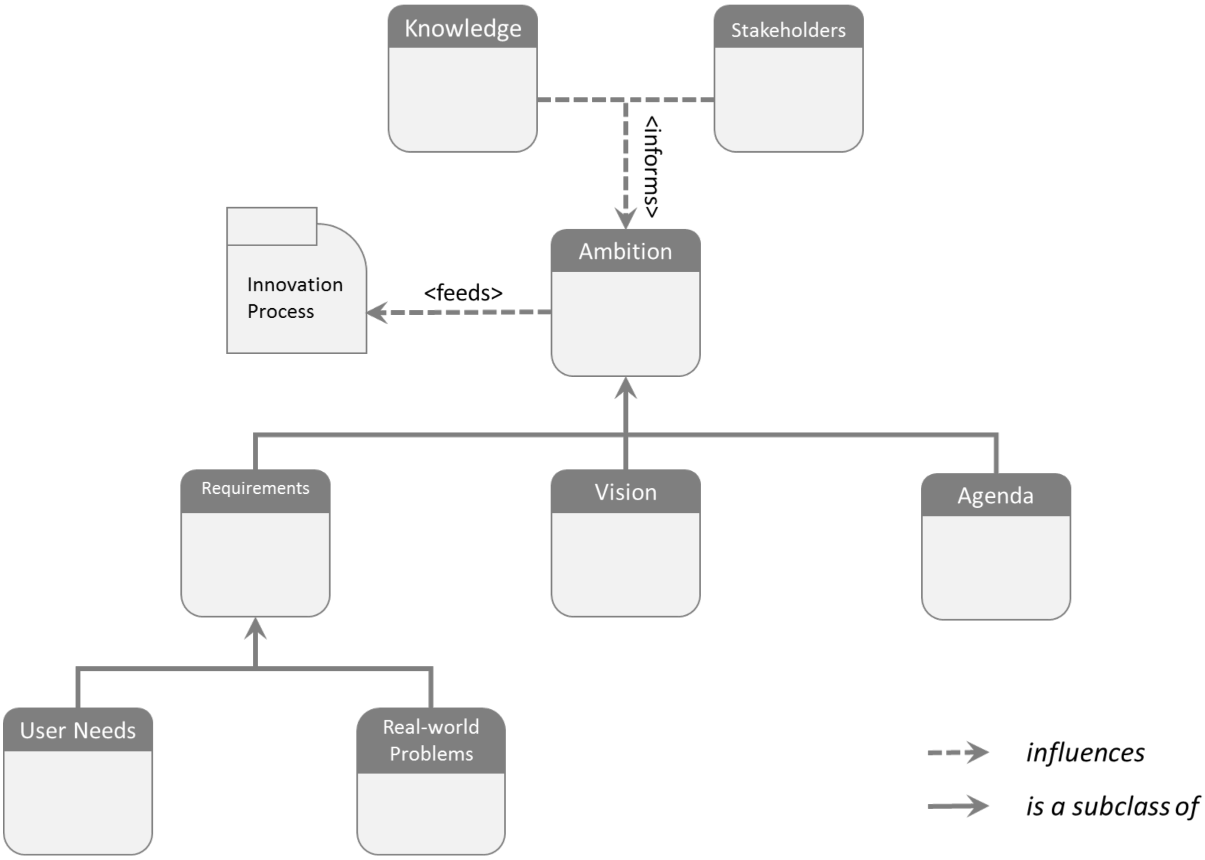


FIGURE 7: HOW INNOVATION AMBITION IS GENERATED

3.5.3.1 Ambition and Innovation

Ambition provides the impetus for innovation as the summary of requirements, justification and / or objectives. It is informed both by *Stakeholders* (see Section 3.5.6) as the sublimation of the views and concerns of all relevant parties, as well as by existing *Knowledge* (Section 3.5.4) which may either constrain or inspire the innovation. *Stakeholders* and *Knowledge* therefore represent the context (social as well as intellectual) within which requirements have arisen and been shaped. *Ambition* as a class is associated with the following subclasses:

- *Requirements* which represents the tangible and near-term goals of *Society,* for example: *automatic braking should be introduced to reduce the effects of driver reaction times*. Such *Requirements* are subdivided into:
    - *User Needs:* issues and challenges recognised by those who operate within the space that the innovation will most likely affect; e.g., *a bicycle that pedals itself uphill*; and
    - *Real-world problems:* issues and challenges which may not be recognised as directly affecting individual users, but which nevertheless represent obstacles to comfort, usability or similar; e.g.: *self-cleaning windows*.
- *Vision* represents the overall possibly strategic aims of *Society*, new ideas and ways of thinking or acting; *Requirements* may be seen as contributing towards this high-level,

Co-funded by the Horizon 2020 Framework Programme of the European Union

long-term target. For example: *society wants all citizens to feel safe wherever they walk at night*; and finally:

- *Agenda* which represents the cultural or intellectual context within which an innovation takes place; e.g., *Society wants to reduce its reliance on fossil fuels*.

Without a suitable *Ambition,* therefore, there would be no need to innovate.

3.5.3.2 Routes to Ambition

An *Ambition* may be initiated in response to any of the following:

- *Stakeholder* input: a perceived need in relation to *Requirements, Vision* and/or *Agenda*;
- *Society*: any area of *Society* (see, for instance, Section 3.5.5) may identify a need and present as near- or long-term *Requirements*, or as part of an overall *Vision* or *Agenda*;
- The *Innovation Process* itself; at any point in the process (*Case, Prototype* or *Outcome*), new ideas or requirements may be generated. These could then potentially seed further investigation or activity.

In short, it is possible to arrive at *Ambition* via all users and the context in which they operate (*Society*), individuals (via the *Stakeholder* construct) or from activity associated with innovation (from the *Innovation Process*).

As a route to innovation, *Ambition* represents the accumulation of inputs from elsewhere in the domain right at the point before innovation is initiated or attempted.

3.5.3.3 Expertise required

Excluding *Knowledge* and *Stakeholders* covered in the following sections, the development and evaluation of *Ambition* is largely down to:

- (ICT) developers and engineers providing the skill to gather, assess and articulate (technical) requirements;
- All *Stakeholders* across all parties with an interest in or affected by a given innovation.

3.5.4 The constructs associated with KNOWLEDGE

Knowledge may be considered the union of all enablers and constraints, other than individual (human) agents. The construct is summarised in Figure 8.

Co-funded by the Horizon 2020
Framework Programme of the European Union

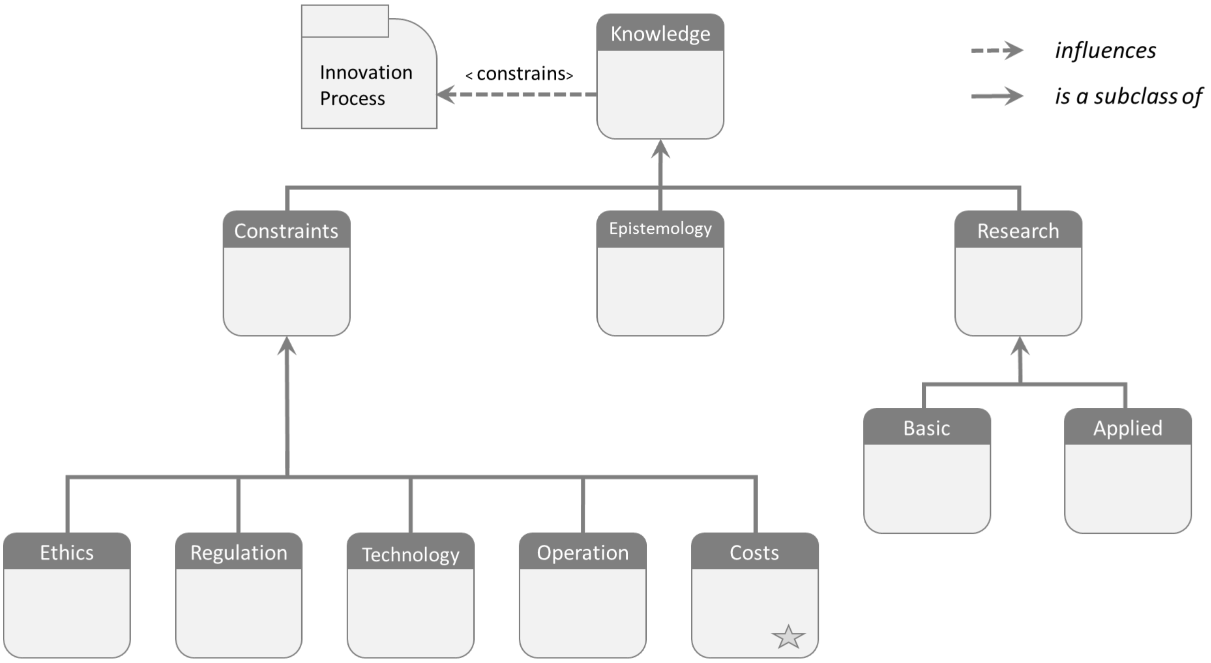


FIGURE 8: KNOWLEDGE AS AN ENABLER AND CONSTRAINT ON INNOVATION

3.5.4.1 Knowledge and Innovation

Knowledge acts as the basic enabler of all innovation: it is on this basis that an innovator is able to move forward and develop a fresh innovation. It sits within the context of all that we do and possibly could know. As enablement, therefore, *Knowledge* in terms of *Research* would be applied to address what is articulated as part of the *Ambition*. But at the same time, *Knowledge* may reflect a range of different constraints, such as physical resource, but also the controls which *Society* imposes. Sub-classes associated with *Knowledge* therefore include:

- *Constraints:* anything which might constrain the *Innovation Process*, which may be the result of one or more of:
  - *Ethics:* a principle which determines appropriate action or behaviours. For example, it would not be ethical to develop a device that creates an unfair advantage of one group over another;
  - *Regulation:* the legal framework within which something would operate; e.g., it is not legal to collect personal data without the knowledge and consent of the data subject;
  - *Technology:* the existing technical or technological background against which something develops. For instance, we are not able to drive vehicles without fuel;
  - *Operation:* the environment within which something would be deployed; for example, it would not be suitable to develop an embedded medical device from corrosive materials;
  - *Costs:* there may be issues of financial or societal cost which limit what can and cannot be done. *NB* this construct is further expanded below (Section 3.5.4.4)
- *Epistemology:* the limitation imposed by what we can know.
- *Research:* this construct relates to the output of human intellectual endeavour; e.g., acquired knowledge. It may be:
  - *Basic:* theoretical investigation which may be independently motivated; the search for knowledge for its own sake;

   - *Applied:* knowledge and experience as derived from trying to address specific and identified issues. For example, investigating the use of certain alloys in dentistry.

Knowledge therefore sets the baseline from which we develop new ideas and turn those ideas into novel solutions or innovations.

3.5.4.2 Contributions to Knowledge

Individuals as well as *Society* as a whole may contribute *Knowledge*. This may include new and extended experience which itself is the result of a previous innovation or innovations in a different domain. *Knowledge* therefore represents the current status of what we know from all our experience.

Knowledge provides a context for innovation, rather than a specific route to it.

3.5.4.3 Expertise required

Apart from more general *Stakeholders* in *Society*, along with ancillary disciples such as *Sociology* and *Psychology*, the primary skills related to *Knowledge* include:

- *Ethics*: to identify any implications for human actors within any resulting network around a given innovation;
- *Legal:* to identify legal or other regulatory issues, such as regulatory approval;
- *ICT:* for their understanding of technology and operational environment;
- *Economics:* to identify the costs (financial or otherwise) associated with a given innovation;
- *Human Geography:* to identify the environmental implications of an innovation; and
- *Philosophy (Epistemology):* to outline the limitations of what we can and cannot know about a given domain.

Experts from these areas should be engaged to support the innovation process.

3.5.4.4 The constructs associated with COSTS

Costs represents a contributory factor to *Knowledge*, mediated by *Constraints*. This construct represents anything which *Society* must expend to enable the innovation; and an important feature of such expense is that once made, it may or may not be easily replaced. This is therefore one area which needs to be considered as part of the business case and sustainability plan for any activity. As such, it is usually part of forming the business case (typically in cost-benefit analysis) and of the development of a business model.

Co-funded by the Horizon 2020
Framework Programme of the European Union

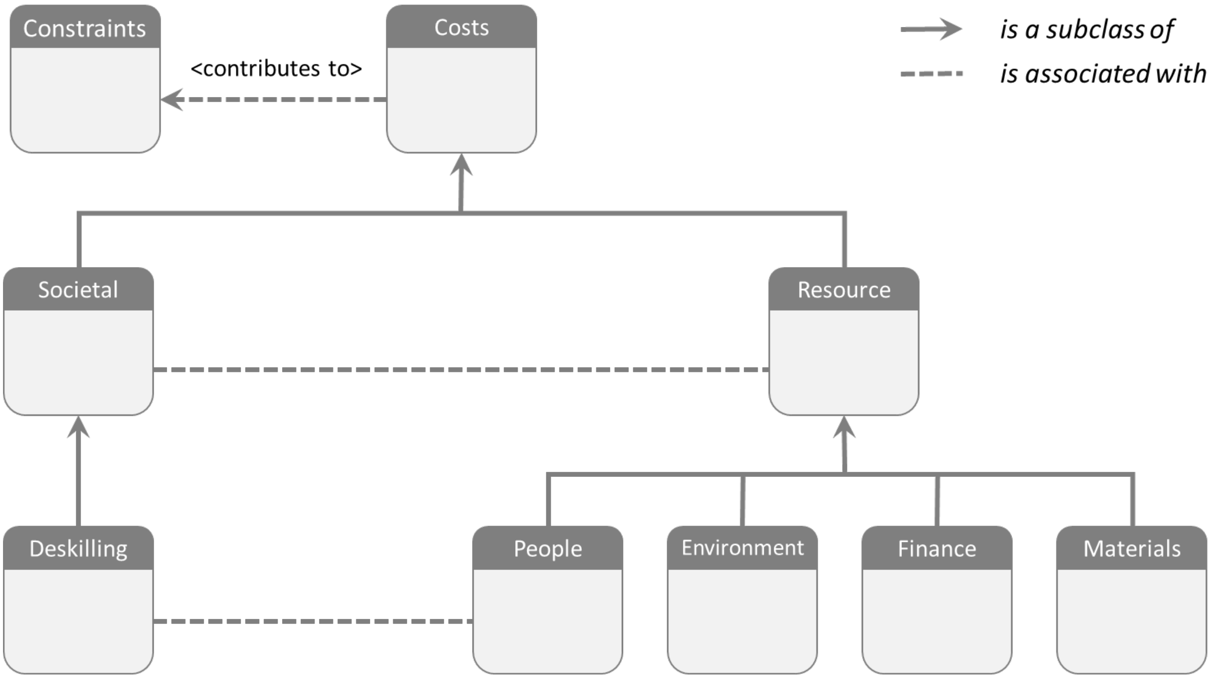


FIGURE 9: SOCIO-ECONOMIC COSTS ASSOCIATED WITH INNOVATION

3.5.4.5 Costs as a Constraint

The most important thing about the *Costs* construct is that it may represent both tangible and intangible expenditure associated with the *Ambition* and the *Innovation Pathway.* Identifying such expenditure is an essential component in developing the *Ambition* especially in terms of exploitation and sustainability, and any associated business plan.

3.5.4.6 Types of Costs

Costs may relate to one of two basic areas:

- *Societal*: covers the negative (socio-economic) implications of a given innovation; that is the effect of a given innovation. It is related to, though not identical with, the more specific
- *Resource*: which summarises more specifically what needs to be available for the *Innovation Process* to run successfully.

The *Societal* class may be extended as follows:

- *Deskilling:* relates to the loss of expertise or skill by those currently engaged in a given task. For example, a neurosurgeon may lack the skill to perform surgery on structures of the mid- or hindbrain as medical technology advances to the extent of allowing keyhole intervention. *Deskilling* is part of the *Societal* implications of an innovation outcome. However, it is also clearly related to the construct *People*.

Expanding *Resource*, we have the following constructs:

- *People:* this relates to the individuals needed to support an innovation, or alternatively those affected by it (see *Deskilling* for instance). For example, the *Innovation process* may only succeed if certain experts are involved, who may not be available and/or who may be expensive to hire;
- *Environment:* the physical context. For instance, if the *Innovation process* requires significant amounts of ground water, this will have an impact;
- *Finance:* this is self-evident.

Co-funded by the Horizon 2020
Framework Programme of the European Union

- *Materials:* what is needed to support the process or exploitation of an innovation outcome. For example, a rare metal.

These all need to be taken into account in developing an innovation.

3.5.4.7 Expertise required

As well as more general *Stakeholders* in *Society*, the primary skills related to *Costs include*:

- *Ethics*: to identify any implications for human actors within any resulting network around a given innovation;
- *Legal:* to identify legal or other regulatory issues, such as approval by a suitable body;
- *ICT:* for their understanding of technology and operational environment;
- *Economics:* to identify the costs (financial or otherwise) associated with a given innovation; and
- *Human Geography:* to identify the environmental implications of an innovation.

Experts from these areas need to be involved in understanding the *cost* implications of an innovation outcome, but also the *Costs* associated with the *Innovation Process* itself.

3.5.5 The constructs associated with SOCIETY

Society is the broad construct which is involved in the initiation of an innovation as well as in taking receipt of the innovation outcome. This is who or what the innovation 'creates value' for (see the definitions in Section 3.2).

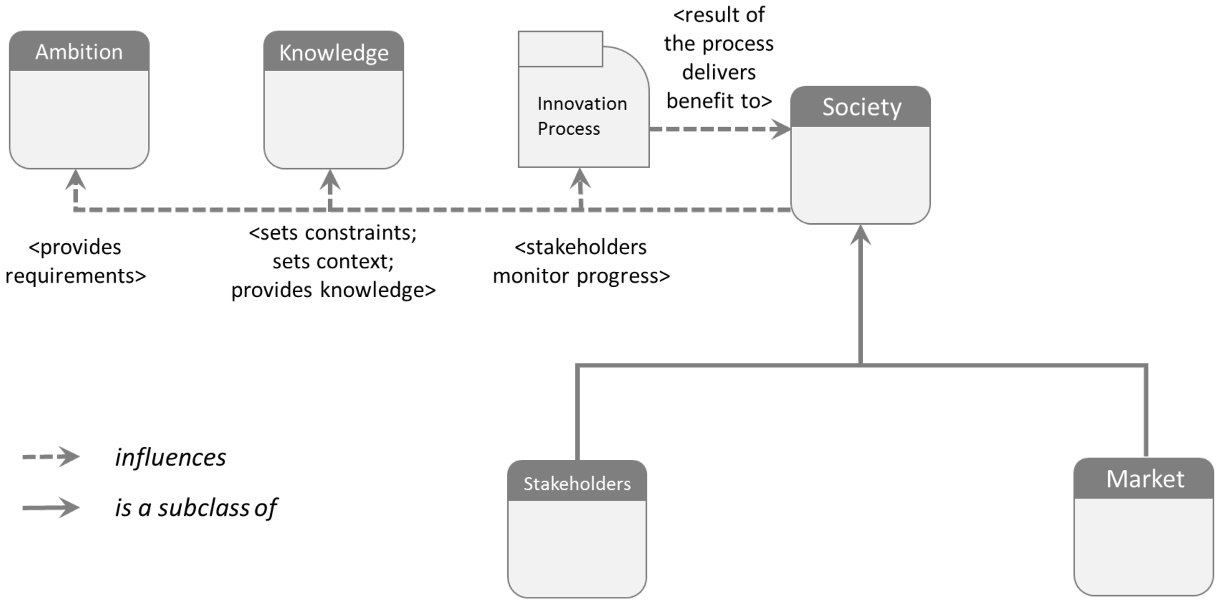

Ambition
Knowledge
Innovation Process
<result of the process delivers benefit to>
Society
<provides requirements>
<sets constraints; sets context; provides knowledge>
<stakeholders monitor progress>
influences
is a subclass of
Stakeholders
Market

FIGURE 10: THE MAKE-UP AND INFLUENCE OF THE CONSTRUCT "SOCIETY"

3.5.5.1 Society and Innovation

The construct *Society* is both influencer and recipient of innovation. *Society* is the context within which *Knowledge* (Section 3.5.4) is created and where new ideas and needs are identified, feeding the definition of *Ambition* (Section 3.5.3), which in turn initiates the *Innovation process* (Section 3.5.1). Once that process has started, then *Society,* in some form or other, will continue to monitor what is going on and what progress is being made.

Co-funded by the Horizon 2020
Framework Programme of the European Union

Once the *Innovation process* completes, the *Outcome* (Section 3.5.1) is delivered back to *Society* to validate its acceptability. This may well lead to more need identification, and then the cycle will repeat, sometimes motivating new innovations.

3.5.5.2 Society as People and Operational Context

Against this background, *Society* comprises two further subclasses:

- *Stakeholders:* these are the people with some interest in innovation *Outcomes*, in terms of providing requirements, but also adding to *Knowledge*, including *Constraints* (Section 3.5.6 below); and
- *Market*: represents the context in which the precursor idea is identified or the operational environment within which the innovation will operate.

So these two constructs will also both seed innovation, proactively through identifying ideas, as well as passively, via observation. In addition, the *Stakeholders* provide checks and balances on the *Innovation process* itself; and, via *Market*, the operational context for innovations is defined.

3.5.5.3 Expertise required

When considering *Society* as a source for needs and ideas to seed innovation, any number of *Stakeholders* may be involved, including ICT skills with experience of handling requirements and developing an appropriate *Ambition*. But specifically with regard to the output of an innovation *Outcome,* monitoring its progress during the *Innovation process*, and then evaluating its acceptability, all skills are required to work collaboratively:

- *Ethics*: to understand any ethical implications of the innovation, and how it may have changed the overall context;
- *Legal*: to position the innovation in regard to the existing regulatory framework;
- *Sociology*: to review how societal structures are affected;
- *Psychology*: to identify the effects on people's behaviours and attitudes;
- *Economics*: to understand the financial and other benefit structures which may accrue because of a given innovation;
- *(Human) Geography*: to evaluate any knock-on effect or environmental consequences of the innovation;
- *(Web) Social Science*: in the longer term, if relevant, to review how behaviour online reacts to the innovation and what this does for data availability.

All such experts may well be engaged during the *Innovation process*, of course, to monitor progress and therefore interact with developers to ensure the ongoing acceptability of what is being done and what is being produced.

3.5.6 The constructs associated with STAKEHOLDERS

As highlighted in the previous section above, *Stakeholders* represent the human actors associated with the *Society*. They may be both recipients of innovation *Outcomes*, and also monitor and add controls to the *Innovation process* itself. For this, there are multiple classes of *Stakeholder* described below.

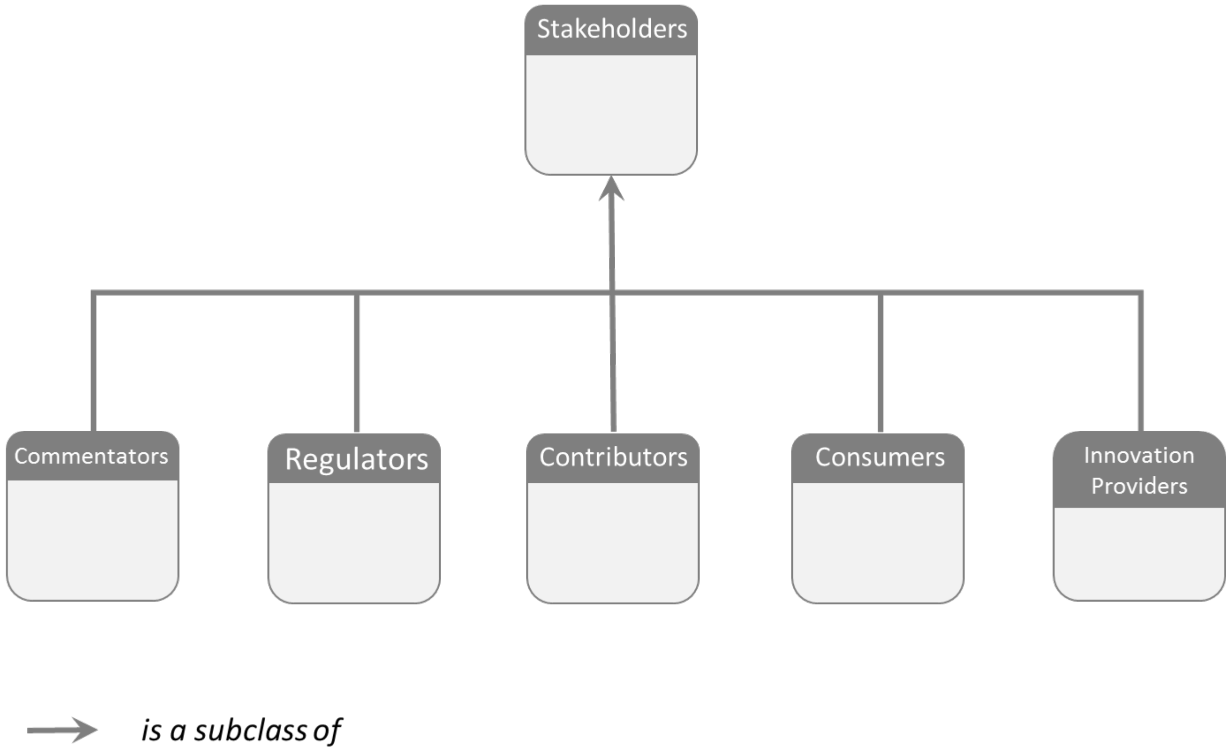


FIGURE 11: THE STAKEHOLDERS IN THE INNOVATION AND INNOVATION PROCESS

3.5.6.1 The Role of Stakeholders

As summarised in Figure 11, the following subclasses ("roles") for *Stakeholders* are defined:

- *Commentators:* any individual or group who monitor innovation, where it is going and what the consequences might be; these *Stakeholders* are responsible for keeping *Society* aware of what is going on, and what may happen (i.e., so that necessary steps can be taken to mitigate risk);
- *Regulators:* those responsible for or interested in all aspects of regulation, including both ethics and legal compliance; these *Stakeholders* would monitor and control any innovation *Outcome*, but also potentially the *Innovation process* itself;
- *Contributors:* these are the *Stakeholders* who have the ideas and identify needs, as well as understand the technologies, in order to be able to help feed the *Ambition* and provide consultancy and expertise during the *Innovation process*;
- *Consumers:* and potential *Prosumers*[61], these *Stakeholders* use the innovation *Outcome*; their role is to evaluate the innovation within a context which works for them[62]; and
- *Innovation Providers:* those *Stakeholders* who run the innovation *Outcome* such that others may access and exploit it.

All of these *Stakeholders* should be involved during the *Innovation process* with a view to maximising the chances of success, and acceptance.

[61] See, for example, https://www.techopedia.com/definition/29581/prosumer which explicitly stresses the difference between a consumer who passively takes on some artefact and a prosumer who may contribute to the development of such an artefact.
[62] The innovation may, for instance, be used for something that was not originally intended.

Co-funded by the Horizon 2020
Framework Programme of the European Union

3.5.6.2 Tussles[63]

The concept of *Tussles* was introduced in the last decade as an attempt to provide a formal model of conflict and its resolution between *Stakeholders*. For example, an online service provider may want to gather, store, process and perhaps even sell personal data and / or indicators of online activity. This would provide them with a commercial advantage and a source of revenue. However, the data subjects may not want their data used for such purposes. They may therefore refuse to use the service. If enough data subjects / users withdraw from the service, then the profits for the provider will reduce and may even collapse. There needs to be some balance, therefore, between users and providers to ensure the fair distribution of advantage and benefit.

This sort of situation will doubtless continue to affect the NGI. In consequence, it is essential that competing expectations and requirements are appropriately managed. This issue will be investigated further as one of the results of the consultation activities described in the opening chapters of this deliverable.

3.5.6.3 Expertise required

All expertise identified (see Section 3.5.2) will need to engage as *Stakeholders* at some time. As a start, the expertise covered in the consultation activities outlined previously will provide an initial indication of the types of skills which would typically be required as *Stakeholders* for innovation.

3.6 NEXT STEPS

In this section of the deliverable, we have outlined an initial domain model for how we envisage the *Innovation process* appropriate for the NGI. In keeping with the white paper on Digital Innovation Networks[64], cross-disciplinary collaboration is essential. In the preceding sections, we have attempted to identify appropriate disciplines to be engaged for specific parts of the ecosystem around the *Innovation process*. The following sections outline how the domain models will be taken forward over the coming period.

3.6.1 Relationship with the Consultation activity

As outlined in the opening chapters of this deliverable, our present activities include a consultation with identified experts in regard to two specific challenge areas as identified in Deliverable D2.1. Although in their initial stages only, they have already begun to inform our work and the first stage domain model we have proposed in this section.

3.6.1.1 Responsible Autonomous Machines

The first round of the Delphi study around *Responsible Autonomous Machines* has now completed and findings from that round are reported above. Already there are two important outcomes from the first round which relate directly to the *Innovation process*.

1. In the more complex areas of the NGI, it is very clear that now, more than ever, appropriate collaboration across disciplines is essential. The message came across from more than one respondent that the complexity of domains such as *Responsible Autonomous Machines* means that we must develop a new paradigm: philosophers and ethicists can no longer remain abstract and academic, they must work directly with the technology developers, at any stage of the innovation development, to define what needs to be taken into account and the contributions expected from any given party. As far as

[63] DOI: 10.1109/TNET.2005.850224

[64] https://www.ict-fire.eu/event/digital-innovation-networks-forum/

our domain model is concerned, this specifically informed the developed of the *Stakeholder* construct with a balance between constraint (checks and balance), consumption, and reportage.

2. Again, a number of respondents reported that as yet we do not even have an unequivocal answer to questions of how and whether technology benefits society. Careful thought ***has to be given*** to issues of who should be allowed to benefit from such advances. They also highlighted elsewhere that with advancing technical capabilities the question of agency needs to be addressed. This prompted a closer look at our construct of *Costs* to include, for example, the notion of *Deskilling* as a potential consequence.

It is clear that the *Innovation process* and our evaluation of innovation pathways needs to consider collaboration and balance as a major criterion.

3.6.1.2 Echo Chambers

The second challenge are identified in Deliverable D2.1 relates to *Echo chambers / Fake News*. Not least against the background of challenging democratic decisions in 2016 across the globe, the reach and power of the Internet has led to a shutdown rather than proliferation of free access to information and opinion[65]. This leads to a perhaps unplanned vulnerability to mechanical or malevolent interference[66]. As this second consultation takes shapes, we need to take on board messages from *Responsible Autonomous Machines* in terms of stakeholders and balanced benefits. This may alter the scope slightly.

3.6.1.3 Tussles

As already mentioned (Section 3.5.6.2) and highlighted once more in the *Responsible Autonomous Machines* consultation, contention between *Stakeholders* and especially between different vested interests in *Society* (see our *Agenda* construct; Section 3.5.3) means that it is now essential to revisit Clark's original *tussle* concept and consider how this should be dealt with for the NGI. This will be developed in moving forward with the prototype tasks in WP3 for instance, and in revising our innovation pathways approach based on this deliverable.

3.6.2 Linking back to WP1 and Impact Evaluation

Table 1 in Deliverable D1.1 proposed a number of different KPIs to evaluate individual initiatives. These metrics may be subdivided into six categories: *Innovation, Economic Sustainability, Technological Maturity, Market Needs,* and *Social Utility.* Within each of these categories, individual measures are proposed. For example, for *Market Needs,* Deliverable D1.1 lists two measures: *Satisfaction of Consumer Market Needs* and *Satisfaction of Enterprise Market Needs.*

As stated already (Section 3.6.1.1 above), the consultation process is beginning to offer slightly different perspectives on the measurement of impact and success. Our *Ambition* construct in the domain model (Section 3.5.3) has been defined in relation to *Agenda* and *Vision*; taken together with the construct of *Stakeholders* we propose and the balance between regulation and consumption, this now suggests that we should revisit the original KPIs with WP1 and consider the following.

[65] The early work by Garrett provides a good starting place to understand some of these effects (Garrett, R.K. (2009) Echo chambers online?: Polticially motivated selective exposure among Internet news users. Journal of Computer-Mediated Communication, 14(2), 265-285) and Pentland's notion of 'Social Physics' in describing online contagion (Pentland, A. (2014) Social physics: How good ideas spread – the lessons from a new science. Penguin)

[66] See for example: Bessi, A. & Ferrara, E. (2016) Social bots distort the 2016 US Presidential election online discussion. *First Monday,* 21(11). Chu, Z., et al. (2010) Who is tweeting on Twitter: human, bot or cyborg? *Proceedings of the 26th annual computer society application conference.* ACM. Howard, P.N. & Kollanyi, B. (2016) Bots, #StrongerIn, and #Brexit: computational propaganda during the UK-EU Referendum. Woolley, S. & Howard, P. (2014) Bad news bots: How civil society can combat automated online propaganda. *TechPresident.*

- As the *Responsible Autonomous Machines* consultation (Round 1) has already highlighted, there needs to be some comparative evaluation of who receives the benefit of technology advance: *Technological Maturity* in terms of KPI category should not be confined to a single group of users; *Social Utility* needs to include multiple perspectives as well. The first task therefore is to review the original set of KPIs as proposed in terms of how they may be interdependent.
- KPIs typically rely on a quantitative assessment (metrics). However, the cross-disciplinary approach which is urgently needed as identified in our first consultation suggests that a corresponding qualitative assessment is needed. The second task in collaboration with WP1 is to review how qualitative assessment may be introduced along with the more traditional quantitative metrics already proposed.

With these potential modifications in mind, one of the first project-internal beneficiaries will be the prototyping activity in WP3.

3.6.3 Relationship with WP3

Notwithstanding any revision as a result of the ongoing consultations in WP2, our initial understanding of innovation pathways based on our domain model will be evaluated within the context of initial prototyping activity in WP3. Within the broader context of existing WP3 plans, we are currently investigating the possibilities described in the following sections.

3.6.3.1 Cross-disciplinary Collaboration

Our domain model suggests the types of expertise, as outlined in the relevant sections above, which need to be involved for a successful innovation pathway based on the *Innovation process* we propose and the associated ecosystem. Within the University of Southampton, we have access to and existing collaboration with experts across all of the fields identified.

To complement the consultation activity, we plan to engage with these experts and task them with designing one or more solutions for the NGI, which may or may not be based on our consultation challenges, and which satisfies the criteria that each individual expert stipulates for their area.

A first-level evaluation of whatever solution(s) they propose will be based on the WP1 KPIs as discussed above.

3.6.3.2 Technical consultation on Challenges

In a second but related activity, we will engage with our own ICT colleagues and ask them to design two solutions to the same NGI problems used for the cross-disciplinary consultation above. The first will be without input from the cross-disciplinary collaborators; the second with that information. Both sets of solutions will again be evaluated against the WP1 KPIs.

Taken as a whole, these activities together will test our innovation pathways within the work package in readiness for more concrete activities in WP3.

4 CONCLUSIONS

This deliverable summarises the current work concerning three related topics following on from D2.1: gap analysis to determine subject areas for further consultation; methods, practice and initial results from two consultations in different subject areas; and how researchers and innovators can be supported to create beneficial and effective solutions to real world applications in the NGI.

The two consultation subject areas selected are “Responsible Autonomous Machines” and “Echo Chambers and Fake News”. The main reasoning for this selection is that they were seen as important, with significant R&D&I potential, but were not yet addressed in detail within the current version of the planned work programme. The first of these is underway and the second will begin in January 2018.

To support innovators, this deliverable has provided an initial investigation into innovation pathways. To avoid ambiguity, a clear distinction is made between the Innovation Process and its result, the Innovation itself, and a model representing these concepts has been proposed. The model will be evaluated through prototyping within WP3, informed by the issues highlighted in the two consultations. Interaction with WP1 is planned, to revisit the KPIs in the light of the issues raised via the consultations.

5 APPENDIX 1 – ETHICAL APPLICATION FOR CONSULTATIONS

This section provides copies of the application to the University of Southampton Ethics Committee for the Responsible Autonomous Machines consultation. The application consists of four documents:

- An application form, describing the study and evaluating the risks (in this case the risks are negligible).
- A Participant Information Sheet, given to potential participants and describing the nature of the study and what will happen during the course of the study.
- A consent form, to be filled in by participants to indicate their consent to take part in the study.
- A Data Protection Plan, describing what personal data will be collected from the participants, and what will be done with it.

Each of these documents is provided in the next four sections. They are forms provided by the University of Southampton, filled in with responses appropriate for the consultation. The Echo Chambers & Fake News consultation will use very similar forms, as the format of the consultation is identical to the Responsible Autonomous Machines consultation.

5.1 FPSE ETHICS COMMITTEE APPLICATION FORM VER 6.6E

Refer to the *Instructions* and to the *Guide* documents for a glossary of the key phrases in **bold** and for an explanation of the information required in each section. The *Templates* document provides some text that may be helpful in preparing some of the required appendices.

Replace the highlighted text with the appropriate information.

Note that the size of the text entry boxes provided on this form does **not** indicate the expected amount of information; instead, refer to the *Instructions* and to the *Guide* documents in providing the complete information required in each section. Do **not** duplicate information from one text box to another. Do not otherwise edit this form.

Reference number: **ERGO/**30743	Submission version: 1	Date: 2017-10-20
Name of **investigator**(s): Steve Taylor		
Name of supervisor(s) (if student **investigator**(s)): N/A		
Title of study: Responsible Autonomous Machines Consultation		
Expected study start date: 2017-11-01	Expected study end date: 2018-02-28	
Note that the dates requested on the “IRGA” form refer to the start and end of *data collection*. These are *not* the same as the start and end dates of the study, above, for which approval is sought. (A study may be considered to end when its final report is submitted.) ***Note*** that ethics approval must be obtained before the expected study start date as given above; retrospective approval cannot be given. ***Note*** that failure to follow the University’s policy on Ethics may lead to disciplinary action concerning Misconduct or a breach of Academic Integrity. By submitting this application, the investigator(s) undertake to:		

Co-funded by the Horizon 2020
Framework Programme of the European Union

- Conduct the study in accordance with University policies governing:
  Ethics (http://www.southampton.ac.uk/ris/policies/ethics.html);
  Data management (http://www.southampton.ac.uk/library/research/researchdata/);
  Health and Safety (http://www.southampton.ac.uk/healthandsafety);
  Academic Integrity (http://www.calendar.soton.ac.uk/sectionIV/academic-integrity-statement.html.
- Ensure the study Reference number ERGO/30743is prominently displayed on all advertising and study materials, and is reported on all media and in all publications;
- Conduct the study in accordance with the information provided in the application, its appendices, and any other documents submitted;
- Submit the study for re-review (as an amendment through ERGO) or seek EC advice if any changes, circumstances, or outcomes materially affect the study or the information given;
- Promptly advise an appropriate authority (Research Governance Office) of any adverse study outcomes (via an adverse event notification through ERGO);
- Submit an end-of-study form if required to do so.

Refer to the Instructions and Guide documents when completing this form and the Templates document when preparing the required appendices.

Pre-study

Characterise the proposed **participants**
Participants are experts in fields related to the subject area of responsible autonomous machines. These are typically autonomous algorithms or applications of AI whose actions need to be explainable and governed from both a legal and ethical standpoint because they are either safety critical or impact the lives of citizens in significant ways. Hence related fields may include but are not limited to artificial intelligence, ethics, law, sociology and computer science in general.

Describe how **participants** will be approached
Experts will be approached using publicly-accessible email addresses. They will be identified using whatever means appropriate, e.g. citation searches, google searches, word of mouth recommendations from partners, etc.

Describe how inclusion and/or exclusion criteria will be applied (if any)
There are no specific criteria, apart from the expertise of the potential participants.

Describe how **participants** will decide whether to take part
Participants will be sent the Information Sheet, the consent form and an additional briefing note. They can read these and if they wish to take part, they can email the investigator, attaching a filled-in consent form.

Participant Information (Appendix (i))

Provide the **Participant Information** in the form that it will be given to **participants** as Appendix (i). All studies must provide **participant information**.

Consent Form/Information (Appendix (iii))

Provide the **Consent Form** (or the request for consent) in the form that it will be given to **participants** as Appendix (iii). All studies must obtain **participant** consent. Some studies may obtain verbal consent (and only present consent information), other studies will require written consent, as explained in the *Instructions, Guide,* and *Templates* documents.

During the study

Describe the study procedures as they will be experienced by the **participant**
The participants will be asked to participate in a remote, non-interactive, anonymous, Delphi Study. All communication will be via email and / or web surveys, so respondents can participate when they have a spare moment. The study consists of three iterations of consultation, with consolidation of the answers in between consultations. The expected timescale for the whole study of three iterations is 3-4 months, beginning Q4 2017. Participation is voluntary and not paid, but low levels of effort are expected from the experts over this timescale: the total estimated effort needed from participants is in the order of 1-2 days, spread over the whole study.

Identify how, when, where, and what kind of data will be recorded (not just the formal research data, but including all other study data such as e-mail addresses and signed consent forms)
Data will be collected by the participants initially emailing textual prose-type responses to the investigator. These will contain their responses to open ended questions regarding what research is needed in to address questions and concerns regarding the subject of the consultation, Responsible Autonomous Machines. The Delphi Method is iterative and later iterations will be in the form of online surveys (most probably using iSolution's online survey tool) where value statements derived from previous iterations of the Study will be presented to participants and participants asked for the strength of agreement / disagreement on a Likert scale. Participant metadata will comprise contact details of participants, their preferences as expressed in their consent forms, as well as the consent forms themselves.

Participant questionnaire/data gathering methods (Appendix (ii))

As Appendix (ii), reproduce any and all **participant** questionnaires or data gathering instruments in the exact forms that they will be given to or experienced by **participants**. If conducting less formal data collection, or data collection that does not involve direct questioning or observation of participants (eg secondary data or "big data"), provide specific information concerning the methods that will be used to obtain the data of the study.

Post-study

Identify how, when, and where data will be stored, processed, and destroyed
If Study Characteristic M.1 applies, provide this information in the **DPA Plan** as Appendix (iv) instead and do *not* provide explanation or information on this matter here.

Study characteristics

(L.1) The study is funded by a commercial organisation: **No** (delete one)

If 'Yes', provide details of the funder or funding agency *here.*

(L.2) There are **restrictions** upon the study: **No** (delete one)

If 'Yes', explain the nature and necessity of the **restrictions** *here.*

(L.3) Access to **participants** is through a third party: **No** (delete one)

If 'Yes', provide evidence of your permission to contact them as Appendix (v). Do *not* provide explanation or information on this matter here.

(M.1) **Personal data** is or *may be collected or processed: **Yes** (delete one)
Data will be processed outside the UK: **No** (delete one)

If 'Yes' to either question, provide the **DPA Plan** as Appendix (iv). Do *not* provide information or explanation on this matter here. Note that using or recording e-mail addresses, telephone numbers, signed consent forms, or similar study-related **personal data** requires M.1 to be "Yes".

(* Secondary data / "big data" may be *de*-anonymised, or may contain **personal data**. If so, answer 'Yes'.)

(M.2) There is **inducement** to **participants**: **No** (delete one)

If 'Yes', explain the nature and necessity of the inducement *here.*

Should participants attend a dissemination event post the Study, it is possible (though not certain yet) that their travel and subsistence costs for attending may be covered.

(M.3) The study is **intrusive**: **No** (delete one)

If 'Yes', provide the **Risk Management Plan,** the **Debrief Plan,** and Technical Details as Appendices (vi), (vii), and (ix), and explain *here* the nature and necessity of the intrusion(s).

(M.4) There is **risk of harm** during the study: **No** (delete one)

If 'Yes', provide the **Risk Management Plan**, the **Contact Information**, the **Debrief Plan**, and Technical Details as Appendices (vi), (vii), (viii), and (ix), and explain *here* the necessity of the risks.

(M.5) The true purpose of the study will be hidden from **participants**: **No** (delete one)
The study involves **deception** of **participants**: **No** (delete one)

If 'Yes' to either question, provide the **Debrief Plan** and Technical Details as Appendices (vii) and (ix), and explain *here* the necessity of the deception.

(M.6) **Participants** may be minors or otherwise have **diminished capacity**: **No** (delete one)

If 'Yes', AND if one or more Study Characteristics in categories M or H applies, provide the **Risk Management Plan,** the **Contact Information**, and Technical Details as Appendices (vi), (vii), & (ix), and explain *here* the special arrangements that will ensure informed consent.

(M.7) **Sensitive data** is collected or processed: **No** (delete one)

If 'Yes', provide the **DPA Plan** and Technical Details as Appendices (iv) and (ix). Do *not* provide explanation or information on this matter here.

(H.1) The study involves: **invasive** equipment, material(s), or process(es); or **participants** who are not able to withdraw at any time and for any reason; or animals; or human tissue; or biological samples: **No** (delete one)

If 'Yes', provide Technical Details and further justifications as Appendices (ix) and (x). Do *not* provide explanation or information on these matters here. Note that the study will require separate approval by the Research Governance Office.

Technical details

If one or more Study Characteristics in categories M.3 to M.7 or H applies, provide the description of the technical details of the experimental or study design, the power calculation(s) which yield the required sample size(s), and how the data will be analysed, as separate appendices.

Appendices (as required)

While it is *preferred* that this information is included here in the application form, it may be provided as separate document files. If provided separately, *name the files precisely* as "Participant Information", "Questionnaire", "Consent Form", "DPA Plan", "Permission to contact", "Risk Management Plan", "Debrief Plan", "Contact Information", and/or "Technical details" as appropriate. Each appendix or document must specify the reference number in the form ERGO/30743/xxxx, the document version number, and its date of last edit.

Appendix (i): **Participant Information** in the form that it will be given to **participants.**

Appendix (ii): Data collection method (eg for secondary data or "big data") / **Participant** Questionnaire in the form that it will be given to **participants.**

Appendix (iii): **Consent Form** (or consent information if no **personal data** is collected) in the form that it will be given to **participants.**

Appendix (iv): **DPA Plan.**

Appendix (v): Evidence of permission to contact (prospective) **participants** through any third party.

Appendix (vi): **Risk Management Plan**.

Appendix (vii): **Debrief Plan**.

Appendix (viii): **Contact Information**.

Appendix (ix): Technical details of the experimental or study design, the power calculation(s) for the required sample size(s), and how the data will be analysed.

Appendix (x): Further details and justifications in the case of: **invasive** equipment, material(s), or process(es); **participants** who are not able to withdraw at any time and for any reason; animals; human tissue; or biological samples.

5.2 PARTICIPANT INFORMATION SHEET

Study Title: **Responsible Autonomous Machines Consultation**

Researcher: Steve Taylor, IT Innovation Centre, University of Southampton S.J.Taylor@soton.ac.uk).

ERGO number: 30743

Please read this information carefully before deciding to take part in this research. It is up to you to decide whether or not to take part. If you are happy to participate you will be asked to confirm your consent via a web form.

What is the research about?

This research is a consultation with domain experts from multiple disciplines to elicit consensus about important research questions, topics and themes in and around the subject area of "Responsible Autonomous Machines". As a definition, Responsible Autonomous Machines are typically autonomous algorithms or applications of AI whose actions need to be explainable and governed from both a legal and ethical standpoint because they are either safety critical or impact the lives of citizens in significant ways, and the consultation's themes strongly correlate with and support those of the current Beneficial AI movement.

The results of the consultation will inform the European Commission on the important research topics surrounding beneficial considerations of AI, and thus assist them to define a future work programme of research within the H2020 framework and FP9. This consultation is part of the H2020-funded HUB4NGI Coordination and Support Action, Grant Agreement No.: 732569, which aims to build communities and contribute to the determination of future research programmes in the context of the next generation of the Internet.

For our consultation it is necessary that consensus should come from the different perspectives offered by experts in multiple disciplines, and we have targeted worldwide experts to invite from disciplines such as AI, machine ethics, asset & threat analysis, the law and computation, explainable AI, the philosophy of computation, the societal impact of the Internet, intelligent machines, psychology and robotics.

A background to the consultation is provided here[67] to give a general idea of some themes. These are to be regarded as starting points only, and any relevant contributions will be welcome. It is a critical objective to elicit relevant research questions, concerns, topics and themes ***not*** mentioned in the background.

Are there any benefits in my taking part?

- The overall benefit to the community at large is the consolidated consensus of opinion from multidisciplinary experts determining recommendations for a research programme into a current hot topic, that of regulation, ethics, responsibility and accountability of autonomous machines and beneficial AI.

Benefits to participants are:

- Participants have a say in influencing new research programmes sponsored by the EC, and to get visible credit for their contribution. The outcome of the consultation is a white paper containing recommendations for research into the key issues in and around Responsible Autonomous Machines. This will be circulated widely across the research community and the EC. The paper will be written by the facilitator, and all

[67] https://www.scribd.com/document/362584972/Responsible-Autonomous-Machines-Consultation-Background-Gateway-Questions

participants who complete the Consultation will be entitled to editorial approval and an author credit.

- We would also like to target a major journal with the consultation results, as it represents the consolidated findings of a consultation into significant concerns regarding the developments in and around AI, crucially from the perspectives of experts in multiple disciplines. Participants completing the whole consultation will be entitled to editorial approval and an author credit on any journal publication outcome from the consultation.
- Participants can benefit from the consultation through seeing alternative perspectives from different disciplines than their own.
- At the end of the consultation, new collaboration opportunities between the participants will be made available (subject to participants' consent).

What is the timescale and estimated effort needed for participation?

The expected timescale for the whole consultation of three iterations is 3-4 months, beginning Q4 2017. Participation is voluntary and not paid, but low levels of effort are expected from the participants over this timescale: the total estimated effort needed from participants is in the order of 1-3 hours, spread over the whole consultation.

Neither travel nor teleconferences are needed. The consultation will take place using the non-interactive mechanism of a web survey, meaning that you can participate at a time that suits you.

Why have I been asked to participate?

You have been asked to participate because of your reputation as an expert in a field relevant to the subject area of Responsible Autonomous Machines. We want this consultation to have high positive impact and genuine substance, so we are targeting the top people worldwide in a mix of disciplines.

What will happen to me if I take part?

You will be asked to participate in a remote, non-interactive, anonymous, consultation that consists of three iterations, with consolidation of the answers in between iterations. This consultation uses the Delphi Method[68], a well-established pattern that aims to determine consensus or highlight differences between expert consultees. This consultation is administrated by a facilitator (Steve Taylor, IT Innovation Centre, University of Southampton, S.J.Taylor@soton.ac.uk) who manages the consultation process and collates results. We expect that three iterations will be undertaken with the overall aim of refining consensus between the responses, Each round will be a separate online survey, and the format of the rounds are described as follows.

1. ***Round 1.*** A selected panel of experts will be invited to participate in Round 1 based on their reputation in a field relevant to the core subject of this consultation. Round 1 is a web survey containing two broad open-ended questions, to which participants can make any responses they wish. It is expected that these responses will be free-form text. Ideally, we would like a side of A4 in total from each participant if possible; though participants are free to submit more if they wish. How ever much participants are able to submit will be used.
2. ***Round 2.*** The participants who completed Round 1 will be invited to participate in Round 2. Using standard qualitative techniques such as grounded theory, the collected corpus of responses from Round 1 will be independently encoded to generate assertions. The assertions are presented back to the participants, and the participants will then be given an opportunity to confirm or revise their opinions in the light of the

[68] Linstone, H.A. and Turoff, M. eds., 1975. The Delphi method: Techniques and applications (Vol. 29). Reading, MA: Addison-Wesley.

consolidated previous results. This will use a structured format web survey (e.g. the participants can agree or disagree with the assertions on a Likert scale).

3. ***Round 3.*** The participants who completed Round 2 will be invited to participate in Round 3. The results of Round 2 will be collated, refining the consensuses and disagreements and assertions will be again generated. These will be presented back to the participants and they are given the opportunity to further confirm or refine their opinions, again using a structured format web survey.

The results of the third round will be collated to determine the final consensus and disagreements.

A key property of the Delphi Method is that it is anonymous during its runtime, in that the participants do not know who the other participants are while the consultation is in progress. The purpose behind this is to avoid halo effects from influential figures within the community. This consultation will be conducted entirely remotely using a series of web surveys, and responses anonymised during the consolidation between rounds. Once the consultation is over, the participants can be revealed to each other[69], so interested participants can collaborate with other participants if they so wish.

Provisional Schedule (all subject to change)

Start Date	Deadline	Days Duration	Activity
01-Nov-17	15-Nov-17	14	Launch of consultation including consent to participate and Round 1 – free text online survey
15-Nov-17	25-Nov-17	10	Facilitator consolidates Round 1 responses
25-Nov-17	09-Dec-17	14	Round 2 – structured online survey derived from Round 1 responses
09-Dec-17	14-Jan-18	36	Facilitator consolidates Round 2 responses (includes Christmas break)
14-Jan-18	28-Jan-18	14	Round 3 – structured online survey derived from Round 2 responses
28-Jan-18	12-Feb-18	15	Facilitator consolidates Round 3 responses & creates first draft of white paper summarising consensus / disagreements
12-Feb-18	27-Feb-18	15	First draft of white paper circulated to participants for comments
27-Feb-18	09-Mar-18	10	Facilitator addresses comments
09-Mar-18	09-Mar-18	0	White paper published

Are there any risks involved?

The consultation is purely a knowledge-gathering exercise, operated remotely over the Internet, so there are no health and safety risks involved.

As a standard part of the Delphi process, all results of the consultation are aggregated so it will not be possible to identify who made what contribution, and there is no risk of any misrepresentation of an individual contributor’s contribution.

[69] Subject to participants’ explicit consent. Those participants that do not wish their identity revealed will be kept confidential.

Will my participation be confidential?

In the outcome, it will not be possible to identify which participant made which contribution. It is a property of the Delphi Method that the facilitator (manager) of the Delphi consultation will aggregate and collate responses from individual contributors, because the method aims to seek consensus between participants.

A key property of the Delphi Method is that it is anonymous during its runtime, in that the participants do not know who the other participants are while the consultation is in progress. Once the consultation is over, the participants can be revealed to each other, so interested participants can collaborate with other participants if they so wish. This is subject to explicit consent from participants, so those participants that do not wish their identity revealed will be kept confidential.

Personal data, in the form of name and email address, will be collected from each participant as mandatory survey questions. This is for the purpose of making sure that only the people that responded to Round 1 get invited to participate in Round 2, and the people that responded to Round 2 get invited to Round 3.

Personal data will be kept until the end of the HUB4NGI CSA project (31 December 2018), and anonymous consultation data will be kept on a secure server behind a firewall for a maximum period of 10 years after the end of the HUB4NGI CSA project.

If explicit consent is given by the participant, their contact details may be retained by the facilitator for the purposes of possible future collaboration.

A consent form is provided in the first round of the survey, where participants can give their consent to participate and also to express their preferences regarding access to and retention of their personal data.

What should I do if I want to take part?

Please fill out the consent questions in the web survey and complete Round 1 of the survey before the Deadline for Round 1 set out in the Schedule above.

What happens if I change my mind?

Anyone can drop out at any time. It is hoped that all participants stay involved until the end of the consultation, but if anyone drops out, there will be no penalty.

Any data collected from participants who drop out will remain in the consultation as anonymous contributions aggregated with other participants'.

The identity of a participant who drops out of the consultation will not be revealed to other participants at the end of the consultation.

Participants who drop out will not be entitled to an author credit, nor editorial approval of, any published outcome of the consultation.

What will happen to the results of the research?

The outcome of the consultation will be a published white paper containing recommendations for research into the key issues in and around Responsible Autonomous Machines. This will be circulated widely across the research community and the EC. We are also targeting publication in a relevant journal for the results of the consultation. Both papers will be written by the facilitator, and all participants who contribute fully to the consultation (i.e. actively contribute throughout the consultation) are entitled to editorial approval plus an author credit for any paper outcome of the consultation (which they have free choice whether to use or not).

All internal data (e.g. individual contributions) will be kept on a secure server behind a firewall. Unless explicit consent for retention has been given in the consent questions, personal data will be kept until the end of the HUB4NGI CSA project (31 December 2018), and anonymous consultation data will be kept on a secure server behind a firewall for a maximum period of 10 years after the end of the HUB4NGI CSA project.

Where can I get more information?

Please contact Steve Taylor, IT Innovation Centre, University of Southampton (S.J.Taylor@soton.ac.uk).

What happens if something goes wrong?

Please contact Southampton University's Research Integrity and Governance Manager (023 8059 5058, rgoinfo@soton.ac.uk).

The University of Southampton has insurance in place to cover its legal liabilities in respect of this consultation.

Thank you.

Thank you for taking the time to read this Information Sheet. It is hoped that you find this consultation interesting, a useful source of ideas and collaborators for your research, and consider it a valuable opportunity to contribute to shaping the European research direction for the next generation of the Internet.

5.3 CONSENT FORM

Study title: **Responsible Autonomous Machines Consultation**

Researcher: **Steve Taylor**
ERGO number: 30743

Please initial the boxes if you agree with the statements below. All are necessary to participate in the Study.

I have read and understood the information sheet for the Responsible Autonomous Machines Consultation (2017-10-20 version 1, ERGO number: 30743) and have had the opportunity to ask questions about the Study.	
I agree to take part in this research project and agree for my data to be used for the purpose of this Study.	
I understand my participation is voluntary and I may withdraw at any time for any reason without my rights being affected.	
I understand my anonymised responses will be included in reports of the Study and possibly additional ethically-approved related research.	
I understand that all participants who participate in the entire Study have editorial approval rights and are entitled to an author credit on the public report that is the outcome of the Study. I also understand that participants who drop out part way through lose these rights, even though their responses may be used anonymously in the public report.	

Optional statements – please only initial the boxes you wish to agree to. None are necessary to participate in the Study – they reflect choices of participants regarding attribution and data retention.

I agree to be named as an author on a public report that is the outcome of this Study.	
I agree to my contact details being revealed to other participants in the Study at the end of the Study.	
I agree to be contacted regarding future unspecified ethically approved research projects or collaborations. I therefore consent to the University retaining my personal details, kept separately from the research data detailed above. I understand that I can request my details be deleted at any time.	

Name of participant (print name)………………………………………………………………………

Signature of participant…………………………………………………………………………………….

Date………………………………………………………………………………………..…………………….

Name of researcher (print name)………………………………………………………………………

Signature of researcher …………………………………………………………………………………….

Date………..

5.4 DPA PLAN

Ethics reference number: **ERGO/**30743	Version: 1	Date: 2017-10-20
Study Title: **Responsible Autonomous Machines Consultation**		
Investigator: **Steve Taylor**		

The following is an exhaustive and complete list of all the data that will be collected (through questionnaires, interviews, extraction from records, etc)

- Publicly-accessible contact details, including name, business or affiliation address, email address, phone number. It is most likely that name and email address will be used, as the Study is conducted over email and via web surveys.
- Consent forms, including name, email address and consent to statements necessary for the study and optional specification of their preferences regarding data retention.
- Responses from participants in the form of textual prose and survey questionnaires.

The data is relevant to the study purposes because survey respondents need to be identified and contacted, and participants' responses are needed. The data is adequate because survey respondents need to be identified and contacted, and participants' responses are needed and the data is not excessive because they need to be contacted, the data is limited to the information necessary to contact them, and participants' responses are needed.

The data will be processed fairly because the participants will have given explicit consent to processing via their consent forms and no other processing will be attempted.

The data's accuracy is ensured because contact details are verifiable and consent forms can be traced back to the original email that came from the data subject.

Data will be stored on a University server and on laptops. The data will be held in accordance with University policy on data retention.

Data files will be protected by secure servers behind firewalls, laptops will be protected by volume level encryption such as BitLocker.

The data will be destroyed by the investigator in accordance with the University's retention policy, unless permission has been given in the consent form for a participant's data to be retained. (e.g. retain contact details so as to enable future collaboration). The data will be destroyed through deletion of files on servers.

The data will be processed in accordance with the rights of the participants because they will have the right to access, correct, and/or withdraw their data at any time and for any reason. Participants will be able to exercise their rights by contacting the investigator (e-mail: sjt@it-innovation.soton.ac.uk).

The data will be anonymised as a fundamental part of the Delphi Method used in this Study. All responses will be kept by the investigator and aggregated so it is not possible to see who said what. The participant's identities will be kept anonymous throughout the course of the Study – again, another fundamental feature of the Delphi Method. Participants can authorise / forbid their name and email to be shared with other participants at the end of the so as to enable collaboration between them. Participants will be offered author credit on a public report that is the outcome of the Study and can either accept or refuse.

All data will be processed inside the European Economic Area (EEA).

Co-funded by the Horizon 2020
Framework Programme of the European Union