TILBURG LAW SCHOOL

LL.M Law and Technology

2021-2022

Master Thesis

ALEXA! Can You Hear Me?

Navigating Trust and Trustworthiness of Virtual Assistant Technologies under the European Union Artificial Intelligence Act.

Ajuwon, A.R.

Anr: 328951

Student Number: 2066541

Supervisor: Dr. Noorman, M.E.

Second Reader: mr.ir. Schellekens, M.H.M.

August 2022

Table of Contents

Chapter I: Introduction

1. Background

Voice Assistant Technologies (VATs) like Google Assistant, Siri, and Alexa have become a part of our lives. Since they are Artificial Intelligence (A.I.), they will be governed by the European Union Artificial Intelligence Act, which is a "Union legislative framework laying out harmonized norms on AI to stimulate the development, usage, and adoption of AI in the internal market that satisfy a high degree of public interest and fundamental rights protection.".[1] Hence the need to look at how the Act impacts and ensures trustworthiness of these systems.

For many, Alexa, Siri, or Google assistant are the first experience with VAT. However, studies into AI-based digital assistants can be traced to ELIZA created in 1966 by Joseph Weizenbaum [2]. VATs have evolved from Apple's Siri in 2010 to Google's voice search, Cortana (Microsoft), Alexa (Amazon), Bixby (Samsung), and the multifaceted Google Duplex, which can book appointments and have real-time conversations on behalf of the owner[3]

Text-based assistants, such as chatbots, have evolved due to technological advancements to meet customers' needs. They are embedded in smart speakers, automobiles, watches, smart televisions, and home appliances. These technologies have been well received, and a considerable number of people interact with them.

Nevertheless, Virtual Assistant applications like Google Assistant, Siri, and Alexa also raise concerns and questions about a breach of trust and ethical standards used in these technologies. Researchers have attempted to measure the trustworthiness of VAT specifically Alexa[4](largest market share) and have found that the third-party applications that are used on Alexa called 'Skills'

[1] European Commission, Proposal for Regulation of the European Parliament and of the Council laying down the harmonized rules on Artificial Intelligence ('AI Act') and amending certain Union legislative acts (Com (2021) 206 final) (Hereafter referred to as the "AI Act").
[2]"The History of Chatbots - from Eliza to Alexa" (*AI Chatbot Platform from Onlim*, December 3, 2021) https://onlim.com/en/the-history-of-chatbots/ . Accessed December 12, 2021.
[3] (Directed by Google Developers YouTube 2018); https://www.youtube.com/watch?v=ogfYd705cRs&gt ; 35.04 - 40:15, Accessed December 12, 2021
[4] Cheng, L. Wilson, C. Liao, S. Young, J. Dong, D and Hu. H. (2020). Dangerous Skills Got Certified: Measuring the Trustworthiness of Skill Certification in Voice Personal Assistant Platforms. Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security. Association for Computing Machinery, New York, NY, USA, 1699–1716. https://doi.org/10.1145/3372297.3423339 Accessed 26th May 2022

have certain limitations in the vetting process by Amazon, proving that a malicious user can publish a Skill under a fake name/brand.[5] They could also make backend code changes after approval to encourage users to reveal unwanted information.[6] There have been varying situations where the trustworthiness of these systems have been brought into question. Examples include in 2018, security researchers at Checkmarx turned an Amazon Echo into a spy device that recorded unsuspecting owners.[7] Amazon Alexa has also been found to say inappropriate content to children instead of playing a song.[8]

In 2018, an Amazon Echo device intercepted a family's private chats and forwarded them to a person on their contact list in Seattle.[9] According to Security Research Labs, there are attack scenarios that apply to both Alexa and Google homes because of weaknesses in these systems, a hacker might phish for sensitive information and eavesdrop on users.[10] In 2019, a security expert in Manchester City was convicted of stalking his estranged wife's home by using the smart device to spy on her.[11]

As proven by the Google Duplex, virtual assistant technologies are advancing at a quick speed, and because these technologies can now interact successfully, they now provide an unprecedented level of fluidity, intelligibility, and autonomy never seen before[12], these technologies are now heavily relied upon in various sectors and by a diverse group of people, necessitating the need to

---

[5] Li B and others, *Advanced Data Mining and Applications 17th International Conference, ADMA 2021, Sydney, NSW, Australia, February 2-4, 2022, Proceedings, Part I* (Springer International Publishing 2022)
[6] Lentzsch, C & Shah, S & Andow, B & Degeling, M & Das, A & Enck, W. (2021). Hey Alexa, is this Skill Safe? Taking a Closer Look at the Alexa Skill Ecosystem. https://10.14722/ndss.2021.23111 Accessed 26th May 2022.
[7] Newman LH, "Turning an Amazon Echo into a Spy Device Only Took Some Clever Coding" (*Wired* April 25, 2018) https://www.wired.com/story/amazon-echo-alexa-skill-spying accessed December 12, 2021
[8] NY Post. (2016). Toddler Asks Amazon's Alexa to Play Song but Gets Porn Instead. https://nypost.com/2016/12/30/toddler-asks-amazons-alexa-to-play-song-but-gets-porn-instead/ Accessed 27th April 2022.
[9] Kim, E. (2018). Echo secretly recorded a family's Conversation and Sent it to a Random Person on their Contact List. CNBC. https://www.cnbc.com/2018/05/24/amazon-echo-recorded-conversation-sent-to-random-person-report.html Accessed 27th April 2022
[10] Braunlein, F & Frerichs, L. (2019). Smart Spies: Alexa and Google Home expose users to vishing and eavesdropping. Security Research Labs. https://www.srlabs.de/bites/smart-spies#:~:text=SRLabs%20research%20found%20two%20possible,assistants%20into%20'Smart%20Spies'. Accessed 27th April 2022.
[11] Burden, E. (2018). Husband used smart-home device to spy on wife. The Times. https://www.thetimes.co.uk/article/husband-used-smart-home-device-to-spy-on-wife-3xzcfqp3m Braitwaite, P. (2018). Smart home tech is being turned into a tool for domestic abuse. WIRED. https://www.wired.co.uk/article/internet-of-things-smart-home-domestic-abuse Accessed 25th April 2022.
[12] Dingli A, Haddod F and Klüver Christina, *Artificial Intelligence in Industry 4.0: A Collection of Innovative Research Case-Studies That Are Reworking the Way We Look at Industry 4.0 Thanks to Artificial Intelligence* (Springer 2022)

evaluate the trustworthiness of these technologies.[13] Security flaws, behavioral monitoring and influence, bias and discrimination, and so on are all examples of vulnerabilities.

VAT systems are used by youngsters, as well as in homes, schools, and businesses, all of which are areas where people spend a significant amount of their time. The gadgets gather data over time, particularly sensitive data, and track user behavior changes, making these technologies high-risk. The amount to which they are used in the house may expose them to additional risks.[14]

Trust is vital for VATs because for the system to be utilized and to work properly, users have to trust the VAT system.[15] Furthermore, trustworthiness of AI is the primary purpose of the AI Act; the EU proposes in the Act that "trustworthy AI delivers advantages and will give consumers the confidence to adopt these technologies while also encouraging enterprises to build trustworthy AI systems."[16] The EU recognizes that AI may give answers to many social concerns; however, this can only be accomplished if the technology is of good quality, produced, and used in ways that gain people' confidence.[17]in the new draft AI act. This raises the question of what it means to improve trust in or trustworthiness of VAT through this new regulation.

The EU is focusing on trustworthy AI rather than another concept because trustworthy AI is based on the idea that trust is the foundation of societies, economies, and sustainable development, and society will only achieve and utilize the full potential of AI if trust can be established in it.[18]

Trustworthy AI is the path taken because other options, specifically ethics and other soft constrai nts, lack a corresponding implementation mechanism and the possibility of ethics washing.[19]

---

[13] Leslie, D. (2019). Understanding Artificial Intelligence Ethics and Safety: A Guide for the Responsible Design and Implementation of AI Systems in the Public Sector. The Alan Turing Institute. https://www.turing.ac.uk/sites/default/files/2019-08/understanding_artificial_intelligence_ethics_and_safety.pdf
[14] Stahl, B.C. (2021). Artificial Intelligence for a Better Future. Springer Briefs in Research and Innovation Governance, Chapter 4. Pp. 35
[15] Bolton T and others, "On the Security and Privacy Challenges of Virtual Assistants" (2021) 21 Sensors 2312
[16] "Excellence and Trust in Artificial Intelligence" (*European Commission - European Commission*May 18, 2022) <https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/excellence-trust-artificial-intelligence_en> accessed July 29, 2022
[17] Ibid
[18] Thiebes S, Lins S and Sunyaev A, "Trustworthy Artificial Intelligence" (2020) 31 Electronic Markets 447
[19] China Academy of Information and Communications Technology (CAICT; 中国信息通信研究院; 中国信通院) and JD Explore Academy, "White Paper on Trustworthy Artificial Intelligence" (2021) https://perma.cc/9XZR-8KNE

From a governance standpoint, trustworthy AI penetrates every core of an enterprise, from opera tions to internal management to research and development.

The High-Level Expert Group on AI (HLEG-AI) defines trustworthy AI as a "foundational goal, with three components: actors and processes involved in AI systems, including their development, deployment, and use, should be ethical, legal, and robust."[20]

Silvia et al. performed a study in 2018 to detect current advancements in VAT and discovered that most research on VAT legislation has concentrated on eight (8) main categories - Education, General, Health, Infrastructure, Privacy Usability, Sports, Education, and Games.[21] The study also claims that there is little research on the ethics and trustworthiness of VAT. As a result, I was inspired to conduct study on trust and VAT.

The role of the AI Act

The European Union's Draft Artificial Intelligence Act offers a possibility to resolve ethical issues concerning emerging technologies such as VATs. The AI Act's purpose is to provide people confidence in AI technologies. It offers a legal framework for trustworthy AI.

With the self-regulation and transparency responsibilities, the AI Act in its current form makes it simple for creators of these systems to carry on as usual. The primary goal of regulation is to affect the behavior of individuals and organizations; but, for regulation to be effective, the system must be capable of holding these entities accountable.

VAT is within the scope of the Act's applicability to this technology, as described in Article 2, which states that "the Regulation applies to suppliers placing on the market or putting into service A.I. systems in the Union... as well as users of A.I. systems situated within the Union."[22] The Act also creates a risk-based framework (prohibited, high and limited risks). VAT is classified as having a limited/medium risk under the Act.

---

[20] Smuha, N.A. [2019] 'The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence' 20(4) Computer Law Review International 97-106. https://doi.org/10.9785/cri-2019-200402 Accessed 28th September 2021

[21] A Silva and others, 'Intelligent Personal Assistants: A Systematic Literature Review' [2020] 147(1) Expert Systems with Applications, https://doi.org/10.1016/j.eswa.2020.113193 Accessed 26th September 2021

[22] European Commission, Proposal for Regulation of the European Parliament and of the Council laying down the harmonized rules on Artificial Intelligence ('AI Act') and amending certain Union legislative acts (Com (2021) 206 final) (Hereafter referred to as the "Act").

This does not imply that such systems are safe or risk-free.[23] VAT poses a significant danger to consumers, and the Act only specifies a transparency need for limited-risk AI and voluntary codes of conduct for providers of these systems. This allows providers to continue doing business as usual while avoiding incredibly important processes that safeguard users of their systems since the law allows them to do so.

How does the definition of trust and trustworthiness in the AI Act apply to VAT? In a general sense, trust means **a steadfast belief or confidence** in the character, ability, strength of someone or something.[24] To comprehend human trust in AI, however, it is vital to investigate it from the perspective of philosophy, psychology, and sociology of how people interact with one another - that is, Interpersonal trust.[25] According to the HLEG-AI consisting of fifty-two (52) subject matter experts, to trust AI means to incorporate explainability, accountability, responsibility, reliability, transparency into intelligent systems.[26] Trustworthy AI is linked to normative assertions about the technology's attributes and often necessitates ethical methods.[27]

Achieving trustworthiness in the technological system, primarily via regulation, seems like an unrealistic goal because the fundamental purpose of the law is to establish standards, maintain order, resolve disputes, and protect liberties and rights, to be effective it must have the ability to hold all organisations and individuals accountable.

2. Research Questions and Sub Questions

The main research question answered by this research is the following:

How does the AI Act's concept of trust and trustworthiness impact Virtual Assistant Technologies, and what steps can be taken to close any gaps?

The research question has been deconstructed into the following sub-questions:

---

[23] Stuurman K and Lachaud E, "Regulating AI. A Label to Complete the Proposed Act on Artificial Intelligence". (2022) 44 Computer Law & Security Review 105657

[24] Merriam Webster Dictionary. https://www.merriam-webster.com/dictionary/trust Accessed 7th December 2021

[25] Jacovi A and others, "Formalizing Trust in Artificial Intelligence" [2021] Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency https://doi.org/10.1145/3442188.3445923 .

[26] Responsible Innovation Project. Can We Trust AI When We Can't Even Trust Ourselves (2021) https://responsibleproject.com/trust-in-ai-can-we-trust-ai-when-we-cant-trust-ourselves/#:~:text=When%20industry%20organizations%20and%20institutions,transparency%20into%20our%20intelligent%20systems. Accessed 14th January 2022.

[27] Toreini E and others, "The Relationship between Trust in AI and Trustworthy Machine Learning Technologies" [2020] Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency

1. What is Artificial Intelligence (AI) and Virtual Assistant Technologies (VAT)?
2. What is trust, and trustworthy AI? And what does it mean in the AI Act?
3. How do VAT systems lead to harms and risks to users and lack of trust in the systems?
4. How does the AI ACT impact trust and trustworthiness of VAT systems?
5. What are the gaps within the Act that impact trustworthiness of VAT systems and what steps can be taken to fill these gaps?

The sub-questions represent the logical steps of the research and the guide through which readers follow the thesis.

3. Literature Review

My research will focus on four (4) major concepts: artificial intelligence (AI), virtual assistant technologies (VAT), trust, and trustworthiness.

The first is artificial intelligence (AI). I used a range of sources to define the notion, including the founding fathers, textbooks, and reports from expert groups and organizations such as the EU High Level Expert Group on AI. The reason for this is that there is no universally recognized definition of AI[28], and it has been used in a few settings both within and outside of the discipline.[29]

Indeed, the number of definitions of AI has expanded as the discipline has progressed and varied throughout time. As a result, there is currently no commonly agreed definition. The many meanings of AI are linked.[30]

The second concept is Virtual Assistant Technologies (VAT). There are several definitions and names for technologies that might qualify as VAT, and I relied on sources that define, distinguish, and state recent VAT developments. First, the European Commission's Digital Transformation Monitor published a study on the topic, which I examined to assess the Commission's stance on VATs.[31]

[28] Ellul, J. Should we regulate Artificial Intelligence or some uses of software? *Discov Artif Intell* **2,** 5 (2022). https://doi.org/10.1007/s44163-022-00021-9
[29] Wang, P. [2019] 'On Defining Artificial Intelligence'. Journal of Artificial General Intelligence 10(2). Pp. 1-37
[30] United Nations Educational, Scientific, and Cultural Organization (UNESCO) and the World Commission on the Ethics of Scientific Knowledge and Technology. 2019. Preliminary Study on the Ethics of Artificial Intelligence. https://unesdoc.unesco.org/ark:/48223/pf0000367823 Accessed 26th January 2022.
[31] European Commission, 'The Rise of Virtual Personal Assistants' 2018, Digital Transformation Monitor, https://ati.ec.europa.eu/sites/default/files/202005/The%20rise%20of%20Virtual%20Personal%20Assistants%20%28v1%29.pdf Accessed on the 18th September 2021.

For the last concepts, trust and trustworthiness in AI and VATs. Researchers have found it difficult to explain these concepts. Research has looked at what trustworthiness means to some extent but has not looked at how to applies to concrete technologies like VATs. According to Bauer, there is a vast range of literature on trust and trustworthiness from many disciplines of study, and there is no agreement on what the notions signify. Bauer claimed that in the subject of organizational trust alone, there are approximately seventy (70) identified definitions of trust.[32] Hardin's work on trust and trustworthiness was evaluated, and he notes that trust is usually mistaken as trustworthiness since they are thought to be the same.[33]

These publications highlight the fact that even the most experienced trust and trustworthiness scholars struggle to define the notions, as well as the challenge of basing regulation on these ideas.

While examining the AI Act, I reviewed various research articles. The first is a study by Veale and Borgesius[34] who examined the proposed AI Act and discovered problems in it. This study is relevant to my research since it explores the interpretation, breadth, and problems of the proposal's ethical concept of "transparency." Second, because of her significant role as the Coordinator of the European Commission's High-Level Expert Group on AI, I evaluated articles[35] by Natalie A Smuha[36]; her paper describes the approach the EU has taken in building a framework for trustworthy AI.

Finally, my study fills a vacuum in the literature since my focus is on VATs and how the current mode in which they work puts their reliability as AI systems into doubt. It investigates how the gaps in the AI Act discourage makers of these technologies from focusing on assuring the trustworthiness of the systems, as well as what the Commission should examine before the Act becomes enforceable.

---

[32] Bauer, P.C. (2019) Conceptualizing Trust and Trust Worthiness, Research Gate. https://www.researchgate.net/publication/262258778_Conceptualizing_Trust_and_Trustworthiness

[33] Hardin, R. Trust and Trustworthiness. (2002) Russell Sage Foundation Publishing.

[34] Veale, M and Borgesius, FZ. 'Demystifying the Draft EU Artificial Intelligence Act' [2021] 22(4) Computer Law Review International. https://arxiv.org/abs/2107.03721v2 Accessed 27th September 2021

[35] Smuha, N.A & Ahmed-Rengers, E & Harken, A & Li, W & MacLaren, J, & Piselli, R & Yeung, K. (2021). How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for An Artificial Intelligence Act. LEADS Lab, Birmingham University. Accessed 28th April 2022.

[36] Smuha, N.A. [2019] 'The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence' 20(4) Computer Law Review International 97-106. https://doi.org/10.9785/cri-2019-200402 Accessed 28th September 2021

4. Methodology and Overview of The Thesis

The goal of this research is to study the shaping of law and new and emerging technologies. This research is inherently a legal study; hence the methodology is based on doctrinal research.

The European Union draft Artificial Intelligence Act is a novel plan by the European Commission to be at the forefront in AI development and to ensure a solid foundation for regulatory implementation and enforcement.[37] This implies that in applying the Act (upon ratification) to a specific technology, there is space for evaluation and interpretation within the boundaries of its provisions.

Chapter two of the thesis addresses the descriptive sub questions (What is Artificial intelligence (AI), Virtual Assistant Technologies (VAT), what is trust, and trustworthiness. These concepts do not have a legal foundation; therefore, the research will be enriched by external concepts and insights from complementary academic disciplines like computer science, mathematics, philosophy, and information technology to answer the questions. The research will also rely on desk review of the technical analysis of these concepts, relying on a broad range of sources like blogs and news websites.

Chapter three answers the research question on how VAT systems lead to harm and risks to users because of the shoddy processes the providers of the systems operate and how trust can be affected. Due to the sheer influence on users' rights and safety, these vulnerabilities have created an inherent lack of confidence in these systems.

Chapter four analyses how VAT will interact with the AI Act. How the current framework upon which these technologies work questions the trustworthiness of these systems and how the gap in the Act further encourages the manufacturers of these systems to continue business as usual without a compulsion to make their systems trustworthy.

Chapter five serves as the concluding chapter that ties everything together and proffers adjustments that can be made before it becomes law.

[37] European Commission, Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM (2021) 206 final) (hereafter 'AI Act').

Chapter II – Definition of Concepts – AI, VAT, Trust, and Trustworthy AI.

1. Introduction

This chapter looks at the core concepts that underpin this thesis from the definition and historical background of Artificial intelligence to what Virtual Assistant Technologies (VATs) are and the different terms by which they are known. The chapter will also investigate the meaning of trust and trustworthiness and the relationship and relevance in AI.

These concepts have been analyzed further by researchers like Jacovi et al[38], and Li et al.[39] wherein they argue that these concepts serve as the foundation upon which AI developers should build systems that can be trusted. An AI model is trustworthy if it can uphold a commitment/task without compromising the user.[40]

2. What is Artificial Intelligence?

One of AI's founding fathers, John McCarthy, gives historical context, fundamental components of AI, and insight into the definition's evolution. *"Artificial intelligence is concerned with approaches for achieving goals in situations where the available information is of a particularly difficult type,"* he explains. The procedures that must be employed are connected to the challenges of the circumstance and are comparable whether the problem solver is human, Martian, or machine.[41]

According to the HLEG-AI, AI is "systems designed by humans that, given a complex goal, act in the physical or digital world by perceiving their environment, interpreting the collected structured or unstructured data, reasoning on the knowledge gained from this data, and deciding the best action(s) to take (according to pre-defined parameters) to achieve the given goal. AI systems can

[38] Jacovi A and others, (2021). "Formalizing Trust in Artificial Intelligence" Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency
[39] Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., Yi, J., & Zhou, B. (2021). Trustworthy AI: From Principles to Practices. *ArXiv, abs/2110.01167*.
[40] Jacovi A and others, (2021). "Formalizing Trust in Artificial Intelligence" Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency
[41] McCarthy, P. 'Mathematical logic in artificial intelligence' [1988] 117(1) Daedalus, Journal of the American Academy of Arts and Sciences 297-311

also be programmed to learn to adapt their behaviour by analysing how their previous actions affect the environment."[42]

Machine learning, a major component of AI, is the capacity of a machine to solve problems autonomously. This is the discipline that is commonly regarded as the core of Artificial Intelligence nowadays. The majority of the enthusiasm around Machine learning is centred on deep learning.[43] Deep learning is a type of machine learning based on layered descriptions of variables known as neural networks[44], and it has made speech-to-text practical on phones and in homes. Its algorithms can be applied to many applications that rely on pattern recognition[45].

Natural Language Processing (NLP) refers to a computer program's ability to comprehend spoken and written human language. It is a branch of machine learning and combines computational linguistics—rule-based modelling of human language—with statistical, machine learning, and deep learning models.[46] According to Vesper and Cohen, NLP is the "computational analysis of linguistic data, most commonly in the form of textual data such as documents or publications".[47] VAT employs of NLP techniques to respond to users' questions in natural language (speech or text) on specific topics or subjects. These questions cover specific topics or subjects consisting of sentences, phrases, and words.[48]

Though AI has been present and researched for decades, recent advances in the subfields of machine learning and deep learning have resulted in numerous opportunities to contribute to the well-being of individuals, the prosperity, advancement of organizations, and societies; however, it

[42] "Ai Hleg Ethics Guidelines for Trustworthy Ai" <https://www.europarl.europa.eu/cmsdata/196377/AI%20HLEG_Ethics%20Guidelines%20for%20Trustworthy%20AI.pdf> accessed June 20, 2022
[43] *Ibid*.
[44] Prabhu, V., Taaffe, K., & Pirrallo, R. (2020). Multi-Layered LSTM for Predicting Physician Stress During an ED Shift. IIE Annual Conference. Proceedings, 1223.
[45] Eric Horvitz, "Defining Ai" (*One Hundred Year Study on Artificial Intelligence (AI100)* <https://ai100.stanford.edu/2016-report/section-i-what-artificial-intelligence/defining-ai> accessed February 12, 2022
[46] Education, I., 2022. *What is Natural Language Processing?* [online] Ibm.com. Available at: <https://www.ibm.com/cloud/learn/natural-language-processing> [Accessed 15 June 2022].
[47] *Verspoor, K. and Cohen, K., 2013. Natural Language Processing. Encyclopedia of Systems Biology, pp.1495-1498.*
[48] *Ibid*

also has a variety of unique ethical, legal, and social challenges that if not handled properly, may severely impede AI's value contributions.[49]

Examples of issues linked with the rapid development and proliferation of AI are manifold[50]. Ranging from risks of invading individuals' privacy (e.g., swapping people's faces in images or videos via Deepfakes[51] or involuntarily monitoring persons using the Clearview AI across the Internet[52], or listening in on private conversations, or the existence of racial and gender bias in AI,[53], Alexa telling a child to put a penny in a socket[54], to the quick and uncontrollable generation of economic losses by autonomous trading proxies (e.g., the loss of millions of dollars due to flaws in high-frequency trading algorithms).[55]

3. What is Virtual Assistant Technology (VAT)?

Voice communication with devices is becoming increasingly popular. VAT is not a large piece of monolithic software; rather, it is made up of a few components, such as speech recognition, understanding, and production.[56]

VAT are software agents that, in response to commands, perform tasks on behalf of a human[57]. They are utilized in a range of applications such as home automation, administration, media playback, and so on, and they analyze human speech and react with synthetic voices.[58]

---

[49] Floridi, L. 2019. Establishing the rules for building trustworthy AI. *Nature Machine Intelligence*, *1*(6), 261–262. https://doi.org/10.1038/s42256-019-0055-y.

[50] Thiebes, S., Lins, S. & Sunyaev, A. Trustworthy artificial intelligence. *Electron Markets* **31,** 447–464 (2021). https://doi.org/10.1007/s12525-020-00441-4

[51] Bloomberg.com. 2022. *Bloomberg - Are you a robot?* [online] Available at: <https://www.bloomberg.com/news/articles/2020-01-06/how-deepfakes-make-disinformation-more-real-than-ever-quicktake#xj4y7vzkg> [Accessed 15 June 2022].

[52] Nytimes.com. 2022. *The Secretive Company That Might End Privacy as We Know it (Published 2020)*. [online] Available at: <https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html> [Accessed 15 June 2022].

[53] Obermeyer, Z., Powers, B., Vogeli, C. and Mullainathan, S., 2019. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), pp.447-453.

[54] BBC News. (2021). Alexa tells 10-year-old girl to touch live plug with penny. https://www.bbc.com/news/technology-59810383 Accessed 5th February 2022.

[55] BBC News. 2022. *High-frequency trading and the $440m mistake*. [online] Available at: <https://www.bbc.com/news/magazine-19214294> [Accessed 15 June 2022].

[56] Pieraccini, R., 2017. *AI assistants*. MIT PRESS. Pp. 22

[57] N. K and others, "Intelligent Personal Assistant - Implementing Voice Commands Enabling Speech Recognition" [2020] 2020 International Conference on System, Computation, Automation and Networking (ICSCAN)

[58] Mutrak, A. Patil, S. Ticlke, A. Nimbalkar, A. Yadav, S. 2021. Intelligent Virtual Assistant – Vision. International Journal for Research in Applied Science & Engineering Technology.

VATs are pieces of software that are created and designed to help users with simple activities, generally by giving information using natural language processing.[59] A virtual assistant technology takes information and sophisticated data from conversations to comprehend and analyse them utilizing advanced Artificial Intelligence (AI), Robotic Process Automation (RPA), Natural Language Processing, and Machine Learning[60]. A VAT is a conversational interface to which you can speak or write in plain language. It is divided into two categories: text-based assistants (chatbots) and voice-based assistants.[61]

Prof. Joseph Weizenbaum created the first natural language processing computer program, ELIZA, in the 1960s. ELIZA was designed to "*show that communication between man and machine was superficial*."[62] ELIZA used pattern matching and substitution methodology in scripted responses to simulate conversation, giving the program the appearance of understanding. Users developed a reaction to Eliza that is now known as the "Eliza Effect - the tendency to unconsciously assume computer behaviours are analogous to human behaviours; that is, *anthropomorphisation*, a phenomenon present in human interactions with virtual assistants".[63]

The resulting milestone in the development of voice recognition technology was achieved in the 1970s at Carnegie Mellon University with "Harpy," which mastered about "1,000 words, a three-year-vocabulary, old's and could understand statements". It could analyse speech that followed pre-programmed vocabulary, pronunciation, and grammatical patterns to decide which word sequences made sense together, decreasing speech recognition mistakes.[64]

---

[59] Webopedia. 2022. *What Is Conversational AI? | Webopedia*. [online] Available at: <https://www.webopedia.com/definitions/conversational-ai/> [Accessed 15 June 2022].
[60] Person, "Top 10 AI-Powered Virtual Assistant Companies" (*AI Magazine*April 26, 2022) <https://aimagazine.com/ai-applications/top-10-ai-powered-virtual-assistant-companies> accessed July 12, 2022
[61] AI Multiple. 2022. [online] Available at: <https://research.aimultiple.com/conversational-ui/> [Accessed 15 June 2022].
[62] Epstein, J; Klinkenberg, W. D (2001). "From Eliza to Internet: a brief history of computerized assessment". *Computers in Human Behavior*. **17** (3): 295–314. doi:10.1016/S0747-5632(01)00004-8. ISSN 0747-5632.
[63] Weizenbaum, J. (1976). *Computer power and human reason: from judgment to calculation*. Oliver Wendell Holmes Library Phillips Academy. San Francisco: W. H. Freeman
[64] Silver S, "A History of Voice Technology" (*Blog*) <https://info.keylimeinteractive.com/history-of-voice-technology> accessed June 16, 2022

Voice assistant technology has advanced since then.[65] Siri, the first contemporary digital virtual assistant put on a smartphone, was debuted as a feature of the iPhone 4S on October 4, 2011.[66] Amazon announced Alexa alongside the Echo in November 2014. Amazon launched a service in April 2017 for creating conversational interfaces for any virtual assistant or interface. Figure 1 depicts a timeline of the evolution of VATs.[67]

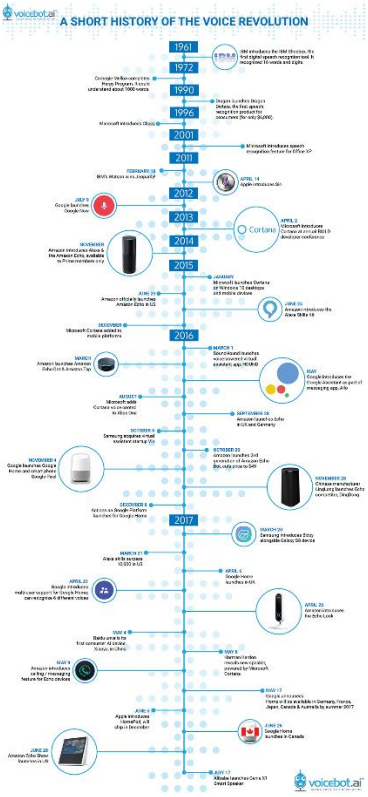

Figure 1: Timeline of Virtual Assistant Technologies.[68]

[65] Ibid
[66]Murph D, "IPhone 4S Hands-on!" (*Engadget*May 13, 2021) <https://www.engadget.com/2011/10/04/iphone-4s-hands-on/> accessed June 16, 2022
[67] Mutchler A, "Voice Assistant Timeline: A Short History of the Voice Revolution" (*Voicebot.ai*March 26, 2021) <https://voicebot.ai/2017/07/14/timeline-voice-assistants-short-history-voice-revolution/> accessed June 16, 2022
[68] Silver S, "A History of Voice Technology" (Blog) <https://info.keylimeinteractive.com/history-of-voice-technology> accessed June 16, 2022

The algorithms may develop data models that identify voice patterns and change them depending on additional data by merging information from the past. Unlike previous varieties such as Eliza, the virtual assistant can answer complicated inquiries, make recommendations and predictions, and even begin a conversation by continually incorporating fresh data about the user's past, preferences, and other user information.[69] Figure 2 depicts a conceptual diagram of VAT.[70]

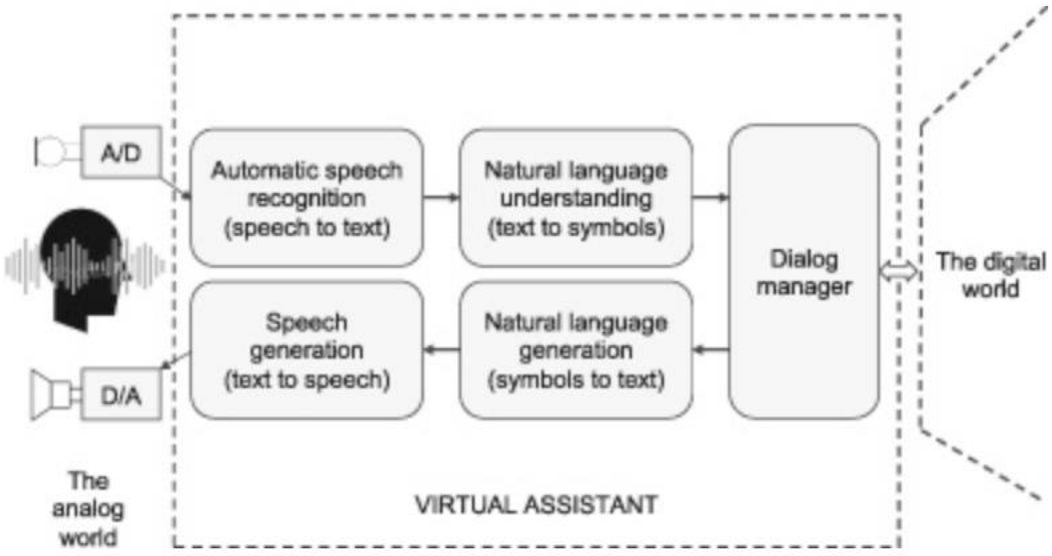

Figure 1: Architecture of a general VAT.[71]

[69] Contact Martijn Kösters Partner mkosters@deloitte.nl, "IPA versus RPA – What's the Difference" (*Deloitte Netherlands* May 6, 2022) <https://www2.deloitte.com/nl/nl/pages/tax/articles/bps-ipa-versus-rpa-whats-the-difference.html> accessed June 16, 2022
[70] Pieraccini, R. (2017) AI Assistants, the MIT Press Essential Knowledge Series, MIT Press. Pp. 22
[71] *Ibid*

While the terminology for VAT that may do activities or offer services for an individual are interchangeable, there are small variances that are mostly decided by how people engage with the technology, the application, or a mix of both.[72]

Here are some examples:

Automated Personal Assistants: Automated means that the task is completed by a machine or software. An automated personal assistant, also known as an Intelligent Personal Assistant, is a "piece of mobile software that can perform tasks or provide services on behalf of a person based on user input, location awareness, and the ability to access information from a variety of online sources (such as weather conditions, traffic congestion, news, stock prices, user schedules, retail prices, etc.").[73] Personal assistants use AI and deep learning to carry out some automated tasks based on the users' experience and behavior with the IPA. They are not required to be conversational.

Intelligent Personal Assistants (IPA): This type of software can help users with simple activities by utilizing natural language. IPAs are intelligent; they can go online and look for answers to a user's questions. A text or voice can initiate an activity. [74]

Voice Assistant: An individual's voice is the most important input. It is a digital assistant that employs voice recognition, speech synthesis, natural language processing (NLP), and artificial intelligence (AI) to deliver service via an application. Siri, Cortana, and Alexa are a few examples.[75]

Smart Assistants: These are often physical gadgets that may deliver a variety of advanced features and services by utilizing smart speakers that listen for a wake-up phrase and can conduct certain activities. A few examples are Amazon Echo, Google Home, and Apple Home Pod.[76]

---

[72] Pandey, A. Vashist, V. Tiwari, P. Sikka, S. Makkar, P. (2020) Smart Voice Based Virtual Personal Assistants with Artificial Intelligence. Artificial and Computational Intelligence, Vol 1, Issue 3.
[73]"Automated Personal Assistant" (*Wikipedia*July 20, 2021) <https://en.wikipedia.org/wiki/Automated_personal_assistant#:~:text=An%20automated%20personal%20assistant%20or,sources%20(such%20as%20weather%20conditions%2C> accessed June 16, 2022
[74] *Ibid.*
[75] Terzopoulos, G. and Satratzemi, M. 2020. Voice Assistants and Smart Speakers in Everyday Life and in Education. Informatics in Education. Vol 19, No 3, 473-490.
[76] *Ibid.*

Chatbots: It does exactly what its name says. It uses text as a medium of communication, as well as to give information and perform activities for users. Chatbots can mimic human-user conversations. Several sectors now use chatbots in their customer support departments. For example, in banking, chatbots may handle accounts and answer simple queries, while in healthcare, patients can utilize chatbots.[77]

There are other names mentioned in literature, which include, digital assistants, intelligent automated assistant,[78] intelligent virtual assistant[79], virtual personal assistant[80], intelligent personal assistant[81], digital butler, digital helper, digital assistant[82], personal digital assistant[83], speech-based natural user interface, voice-activated intelligent personal assistant[84], virtual agent-based daily assistant[85], algorithmic assistant[86], etc.

This thesis employs Virtual Assistant Technology as an umbrella term to encompass all the various agents that exist. Using a single word such as IPA, VPA, IVA, and so on does not cover all aspects because some of these agents interact and are sometimes intertwined. For example, an Amazon Echo device has Alexa incorporated in it and can assist the user with various tasks. VAT is an all-encompassing phrase because it encompasses technologies, whether hardware or software, and it is not confined to items that can be used in a private capacity, as these technologies are now being employed in more public facing sectors such as medicine, education, corporate, and so on. It is

[77] *Ibid.*
[78] "Intelligent Automation Assistant - Patent HK-1220023-A1 - Pubchem" (*National Center for Biotechnology Information. PubChem Compound Database*) <https://pubchem.ncbi.nlm.nih.gov/patent/HK-1220023-A1> accessed June 16, 2022.
[79] Lamontagne, L. Laviolette, F. Khoury, R. Bergen-Guyard, A. 2014. A Framework for Building Adaptive Intelligent Virtual Agents, 15h IASTED International Conference on AI and Applications, pp. 17-19
[80] Tur G, Deoras A and Hakkani-Tür D, "Detecting out-of-Domain Utterances Addressed to a Virtual Personal Assistant" [2014] Interspeech 2014 pp. 283 -287.
[81] Canbek, N.G. Mutlu, M.E. .2016. On the Track of AI Learning with Intelligent Personal Assistant, Journal of Human Sciences, Vol 13 No 1, pp. 592-601
[82] Stucke ME and Ezrachi A, "How Your Digital Helper May Undermine Your Welfare, and Our Democracy" [2017] SSRN Electronic Journal.
[83] Geiger C and others, "Testable Design Representations for Mobile Augmented Reality Authoring" Proceedings. International Symposium on Mixed and Augmented Reality, pp. 145-146.
[84] Lopes, G. Quesada, L. Guerrero, L.A. 2018 Alexa vs Siri, Cortana vs Google Assistant: A Comparison of Speech Based Natural User Interface, International Conference on Applied Human Factors and Ergonomics, Springer, 241-250.
[85] Yaghoubzadeh, K. Kramer, M. Pitsch, K. Kopp, S. 2013, Virtual Agents as Daily Assistants for Elderly or Cognitively Impaired People, International Workshop on Intelligent Virtual Assistants, Springer, Berlin. PP. 79-91.
[86] Gal, M.S. (2018) Algorithmic Challenges to Autonomous Choice, Michigan Telecoms and Tech Law Review.

also worth noting that the European Commission regards these technologies as VAT, both are conversational and employ NLP.[87]

Between 2021 and 2027, the Conversational AI Market in Europe is predicted to develop at a compound annual growth rate (CAGR) of 15.9 percent.[88] During the projected period, the increase in demand for AI-based support services is likely to move the conversational AI market ahead.[89] Furthermore, the increased deployment of omnichannel (seamless and frictionless, high-quality customer interactions that occur within and between contact channels) approaches is predicted to boost the growth of the conversational AI industry.

Google, Apple, Microsoft, Amazon, and Meta (Facebook) are the five digital titans that give their own intelligent personal assistants - VATs like as Siri, Now, Cortana, Alexa, and M - with which AI has provided communication and engagement in recent years.[90]

4. What is Trust and Trustworthiness?

For years, the Commission has aided and enhanced cooperation on AI across the EU to boost competitiveness and certify trust based on EU values "such as promoting scientific and technological progress, combating social exclusion and discrimination, and promoting peace, among others."[91]

Following the launch of the European AI Strategy in 2018 and considerable stakeholder engagement, the High-Level Expert Group on Artificial Intelligence (HLEG) established Guidelines for Trustworthy AI in 2019, followed by an Assessment List for Trustworthy AI in 2020. The first Coordinated Plan on AI was launched in December 2018 as a shared commitment with Member States. The Commission's White Paper on AI, released in 2020, articulated a clear vision for AI in Europe: an ecosystem of excellence and trust, setting the framework for the

[87] (*Advanced Technologies for Industry*) <https://ati.ec.europa.eu/> accessed June 16, 2022
[88] Research and Markets, "European Conversational AI Market 2021 - 2027: German Market Dominated in 2020 and Is Forecast to Reach $202.8 Million by 2027" (*GlobeNewswire News Room*December 13, 2021) <https://www.globenewswire.com/news-release/2021/12/13/2350501/28124/en/European-Conversational-AI-Market-2021-2027-German-Market-Dominated-in-2020-and-is-Forecast-to-Reach-202-8-million-by-2027.html> accessed July 12, 2022
[89] *Ibid.*
[90]Techcrunch.com. 2022. *TechCrunch is part of the Yahoo family of brands*. [online] Available at: <https://techcrunch.com/gallery/a-battle-royale-of-digital-assistants-the-big-5/> [Accessed 15 June 2022].
[91]European Commission. Europe fit for the Digital Age: Commission proposes new rules and actions for excellence and trust in Artificial Intelligence. 2021. https://ec.europa.eu/commission/presscorner/detail/en/IP_21_1682 Accessed 23rd February 2022.

proposal. The European Commission High-Level Expert Group on AI (HLEGAI) adopted the Ethics Guidelines for Trustworthy AI in April 2019, emphasizing that humans will only be able to reap the full benefits of AI if the technology can be trusted. AI that is trustworthy is ethical, legal, and robust.[92]

The European Commission presented a proposal for an artificial intelligence regulation – the AI Act – on April 21, 2021. The draft Act aims to establish harmonised rules for the development, market placement, and use of AI systems that may differ in characteristics and risk, includes bans and an EU product safety conformity evaluation system.[93]

In the explanatory memorandum, the AI. Act expressly aspires to build a trust ecosystem by offering a legal framework for trustworthy A.I.[94] The word "trust" appears several times. The word Trustworthy is mentioned 21 times, Trust 14 times, entrusted 6 times, Trustworthiness 3 times, trustful is mentioned 2 times, and the words Trusted, and Entrusting are mentioned 1 time.[95] Meaning, trust appears to be an essential aspect concerning A.I. and a key component of the Act. The governance system (i.e., institutions and processes meant to assure accountability, openness, the rule of law, and broad-based involvement) and the regulators who use it appear to require public confidence.[96]

It is necessary to define trust when discussing trustworthy AI. Simply expressed, trust implies thinking that someone is honest and will not hurt you, or that something is secure and dependable.[97] It is defined by Webster Dictionary as "certain reliance on the character, ability, strength, or truth of someone or something."[98] However, various social science disciplines have conducted extensive research on the concepts of trust and trustworthiness and anybody interested in the issue will come

[92] "Ai Hleg Ethics Guidelines for Trustworthy Ai" <https://www.europarl.europa.eu/cmsdata/196377/AI%20HLEG_Ethics%20Guidelines%20for%20Trustworthy%20AI.pdf> accessed June 20, 2022
[93] Veale M and Zuiderveen Borgesius F, "Demystifying the Draft EU Artificial Intelligence Act — Analysing the Good, the Bad, and the Unclear Elements of the Proposed Approach" (2021) 22 Computer Law Review International 97
[94] Tschopp PQ-Rand M, "Relationship between Trust and Law Is Counterintuitive and Paradox" (Can Laws build Trust in AI?), https://www.scip.ch/en/?labs.20210916&gt ; accessed December 12, 2021
[95] Tschopp PQ-Rand M, "Relationship between Trust and Law Is Counterintuitive and Paradox" (Can Laws build Trust in AI?), https://www.scip.ch/en/?labs.20210916&gt ; accessed December 12, 2021
[96] Ibid
[97] Cambridge Dictionary https://dictionary.cambridge.org/dictionary/english/trust Accessed 20th February 2022
[98] Webster Dictionary https://www.merriam-webster.com/dictionary/trust Accessed 20th February 2022.

across a large number of - often unclear - definitions.[99] It has also led in a number of important attempts to categorize trust definitions and studies.[100] Bauer compiled a list of the twelve (12) most influential definitions of Trust during his research. However, the number found in literature is higher.[101] In the subject of organizational trust alone, there are approximately seventy acknowledged definitions of trust.[102] Scholars have coined several different sub concepts to supplement this striking abstract complexity, such as "particularized trust" or "knowledge-based trust.".

For the purposes of this study, a composite definition of trust is proposed, based on definitions from prominent papers: "trust is the readiness of one party to be susceptible to the acts of another party; when trust is present, risk-taking behaviour may be observed.".[103] In the face of the difficulties in defining trust, there may be a nucleus common to most of them, which can be identified in the concept of expectation. For example, x believes that y inhabits some role, and that y will competently perform the acts associated with that role. For example, x believes in his doctor or his car mechanic.[104]

According to Luhmann, trust is often viewed as a key approach for coping with the inherent unpredictability of social existence.[105] When analysing trust, there are a few key concepts that must be considered[106] and they are;

a. There must be two players/actors—a party who expects the other actor to perform positively (trustee).

[99] Bauer, P.C. (2019) Conceptualizing Trust and Trust Worthiness, Research Gate. https://www.researchgate.net/publication/262258778_Conceptualizing_Trust_and_Trustworthiness
[100] Lewicki RJ and Bunker BB, "Developing and Maintaining Trust in Work Relationships" [1996] Trust in Organizations: Frontiers of Theory and Research 114." In Conflict, Cooperation, and Justice: Essays Inspired by the Work of Morton Deutsch, edited by Barbara B Bunker and Jeffrey Z Rubin. San Francisco, CA: Jossey-Bass. Bromiley, P., Cummings, L.L. (1995). "Transactions Costs in Organizations with Trust." Research on Negotiation in Organizations 5: 219–50.
[101] Bauer PC, 2013. "Clearing the Jungle: Conceptualizing and Measuring Trust and Trustworthiness" SSRN Electronic Journal
[102] Seppanen, R. Bloomqvist, K. Sundquvist, S. (2007) Measuring Inter-Organizational Trust – A Critical Review of the Empirical Research in 1990-2003. Industrial Marketing Management 36 (2) 249 -258.
[103] Mayer RC, James DH and David SF, "An Integrative Model of Organizational Trust - JSTOR" (July 1995) <https://www.jstor.org/stable/258792> accessed June 16, 2022 33. Pp. 225-232.
[104] Ibid.
[105] Luhmann, N. 1979. Trust and power. Two works by Niklas Luhmann. Translated by Howard Davis. New York: John Wiley & Sons Ltd. [Google Scholar]
[106] Möllering Guido, *Trust: Reason, Routine, Reflexivity* (Emerald 2008)

b. Something must be “at stake” for the trustor. Trust is dangerous and inextricably linked to vulnerability.[107] If there is nothing to lose, trust is unnecessary.
c. Trust cannot be forced. This is largely due to the trustor's dependence on the acts of the trustee. For trust to exist, the trustee must be granted authority.[108]

Trust is intrinsically connected to uncertainty since it cannot be enforced. When it is feasible to foretell what will happen in the future, trust loses its significance.[109] Rather than seeking to decrease uncertainty, trust is a constructive acknowledgment of human life's unpredictability.[110] Trust necessitates a leap of faith.

This leads us to Mayer et al 's ABI framework which states that the three major traits that will impact the appraisal of a party's trust are Ability, Benevolence, and Integrity.[111] A party's ability is the collection of skills, competencies and qualities that allows them to exert influences in a certain domain. Benevolence refers to the extent to which a trustee is thought to want to do good for the trustor. Integrity means the trust between parties is founded on the trustor's conviction that the trustee adheres to a set of standards that the trustor considers acceptable.[112]

Dietz and Den Hartog modified this framework to create the ABI+ model, which stands for Ability, Benevolence, Integrity, and Predictability.[113] The other model prioritizes integrity and competence on alongside predictability. However, there are compelling reasons to include predictability and dependability in the model.[114] Predictability denotes consistency and regularity of behaviour, and it differs much from integrity or competency.[115]

[107] Baier A, (1986). “Trust and Antitrust” 96 Ethics 231
[108] Keymolen E and Van der Hof S, (2019. “Can I Still Trust You, My Dear Doll? A Philosophical and Legal Exploration of Smart Toys and Trust” 4 Journal of Cyber Policy 143
[109] Luhmann, N. 1979. Trust and power. Two works by Niklas Luhmann. Translated by Howard Davis. New York: John Wiley & sons Ltd.
[110] Keymolen E, “Trust in the Networked Era” (2018) 22 Techné: Research in Philosophy and Technology 51
[111] Mayer-Schönberger, Viktor and Cukier K, *Big Data: A Revolution That Will Transform How We Live, Work and Think* (John Murray 2017)
[112] Mayer RC, James DH and David SF, “An Integrative Model of Organizational Trust - JSTOR” (July 1995) <https://www.jstor.org/stable/258792> accessed June 16, 2022 33. Pp. 225-232
[113] Dietz G and Den Hartog DN, “Measuring Trust inside Organisations” (2006) 35 Personnel Review 557
[114] Mishra, A.K. (1996), “Organizational responses to crisis: the centrality of trust”, in Kramer, R.M. and Tyler, T.R. (Eds), Trust in Organisations: Frontiers of Theory and Research, Sage, Thousand Oaks, CA, pp. 261-87
[115] Dietz G and Den Hartog DN, “Measuring Trust inside Organisations” (2006) 35 Personnel Review 557

Predictability is distinct from integrity or competence in that it refers to behaviour consistency and regularity.[116] Predictability's role emphasizes the significance of retaining trust over time through recurring connections. Although they are connected and may reinforce one another, these four characteristics are distinct. Even if one party perceives that one of the traits is lacking, one party can still trust the other.[117]

People are more inclined to trust contemporary technology when it is supplied by a high-reputation institution - signifying skill, benevolence, honesty, and predictability - than when it is offered by a low-reputation organization.[118] Trust in chatbots, for example, is determined by the perceived security and privacy of the provider.[119] Furthermore, trust in technology grows when it is perceived to be reliable, transparent, and secure.[120] [121]

Trust, on the other hand, is a reaction to the technologies created or the procedures by which they were developed. They may not always be ethical. For example, the system might be intentionally prejudiced. Users may not trust technology if it is prejudiced. The ethical problems that underpin the adoption of an AI-based product or service may have an impact on trust perception, for example, if trust depends on having faith in the service not to record conversations while switched off, but the system is built to collect data by continually listening. Users' faith in the system is impacted when they are aware of the device's covert recording.

The ABI+ model best describes the trust relationship between a user's trust relationship with and VAT technology. Siri, for example, can tell the date, weather, and so on when asked, and there is an expectation that these systems are intended to be beneficial and positive for users. There is also the expectation that VAT systems are designed following societally acceptable principles.

5. Trustworthiness

[116] Ibid
[117] Mayer RC, James DH, and David SF, 1995. "An Integrative Model of Organizational Trust - JSTOR" <https://www.jstor.org/stable/258792> accessed June 16, 2022, 33. Pp. 225-232
[118] Siau, K., & Wang, W. (2018) Building Trust in Artificial Intelligence, Machine Learning, and Robotics. *Cutter Business Technology Journal*, 31(2), pp. 47-53.
[119] Følstad A, Nordheim CB and Bjørkli CA, "What Makes Users Trust a Chatbot for Customer Service? an Exploratory Interview Study" [2018] Internet Science 194
[120] Hancock, P.A., Kessler, T.T., Kaplan, A.D., Brill, J.C., & Szalma, J.L. (2020) Evolving Trust in Robots: Specification Through Sequential and Comparative Meta-Analyses. *Human Factors*, pp. 18720820922080-18720820922080. https://doi.org/10.1177/0018720820922080
[121] Hauk N and Hauswirth M, 1970. "Navigating through Changes of a Digital World" (*SpringerLink* January 1,) <https://link.springer.com/chapter/10.1007/978-3-030-86144-5_37> accessed June 20, 2022

It is important to draw a distinction between trust and trustworthiness.[122] Trustworthiness is an attribute, a feature shared by humans/animals or AI systems, and it influences whether others choose to trust us and what occurs next. It is founded on characteristics, beliefs, goals, intentions, competences, and so on, and it is difficult to demonstrate to others.[123]

A trustee might be trustworthy, e.g., never lie, regardless of the trustor's level of trust in him, in fact, irrespective of whether he is trusted by anybody.[124] There is a complexity that is associated with defining trust but this, according to the philosopher Russell, is derived from the intricacy of trustworthiness. In a sense, trusting someone in some context can be described as merely the expectation that the person is going to be trustworthy.[125] The ABI + theory is relevant here as predictability is key to achieving trustworthiness. Trustworthiness is a motivation or a set of motivations for acting. It is very frequent to misconstrue trust for trustworthiness because there is the assumption that they mean same thing. However, even researchers have it difficult distinguishing the two terms.

If discussions of trust are to be comprehended, it is pertinent to specify more narrowly how the term is meant to be used and this applies to how the term is used and applied in relation to artificial intelligence precisely Virtual Assistant Technologies.

6. What is the relationship and relevance of trust and trustworthiness to AI?

All examples of AI in use today may be classified as 'narrow' AI. They are conceived of, created for, and assessed based on their capacity to execute specified tasks within a narrowly defined area – this is how VATs are designed to work[126]. The inability of limited AI to generalize its 'intelligence' to other areas is one of its distinguishing features. A virtual voice assistant, for example, may be trustworthy at understanding spoken dates and converting them into online calendar entries, replying to questions about how to resolve discrepancies, or recommending places or individuals to invite. Such a system, however, would fail badly at reading an X-ray or driving a car. Similarly, a doctor cannot expect to bring the software underlying her autonomous

[122] Freitas, R. and Iacono, S., 2021. *Trust Matters*. London: Bloomsbury Publishing Plc.
[123] *Ibid*
[124] Levi, M and Stoker, L, "Political Trust and Trustworthiness" (2000) 3 Annual Review of Political Science 475
[125] Hardin R, *Trust and Trustworthiness* (Russell Sage Foundation 2002)
[126] Lewis, P.R. Marsh, S. *'What is it like to trust a rock: A functionalist perspective on trust and trustworthiness in AI?'* (2022) Cognitive Systems Research. Vol 72, pg.33-49.

automobile into the office to aid analyse X-ray/MRI data. Not every narrow AI resembles an agent. Many AI-based tools are now integrated into larger software systems, with the decision or suggestion being only one aspect of the tool's use.

What is it like to have faith in these AI-powered tools? It depends on who the trustor is and what they intend to trust it for. It will be like having a well-founded notion that AI technology would be far superior to what is anticipated of a person, but then seeing that this is only true in a certain and small domain.[127] For example, I may feel that an AI-based chess opponent is a better chess player than myself or any person alive today. Similarly, a doctor may assume that a system's ability to offer an accurate MRI analysis exceeds their own.[128] However, the doctor may consider that the system's ability to recognize the ramifications of delivering an inaccurate diagnosis or giving sympathetic treatment to a patient is lacking.[129] It's typically like believing that if I know the system's working constraints, it can accomplish what I need it to do. It's as if you find it mostly predictable, yet occasionally find its behaviour strange and wonder why it did what it did. When trying to do anything, it might feel like you've reached the limits of its capabilities, which can be both irritating and amusing. Trusting AI arises from the notion that when the system accepts a command, the user and the system have reached an agreement. Knowing that the system will fulfil the instruction, but also realizing that it may fail due to reasons outside its control, such as environmental conditions or a lack of internet connection. It might also suggest that the machine has other goals that override it, such as a safety mode that engages the charging cycle or a goal specified by the manufacturer to try to modify users' behaviour.[130]

We must be careful with our words: do we mean to have trustworthy AI because we believe AI can be trusted?[131] In the sense that non-human entities can be assigned trust dynamics in interpersonal relationships? Do we want to build capable, benevolent, trustworthy, and predictable AI systems, or do we want to build AI systems on which humans may rely without necessarily trusting? Because a user can rely on a VAT system to tell the weather in Tilburg or turn off lights

[127] Ibid
[128] *Ibid*
[129] Korot E., Wagner S.K., Faes L., Liu X., Huemer J., Ferraz D., Keane P.A., Balaskas K. 'Will AI replace ophthalmologists?' (2020) Translational Vision Science & Technology, 9 (2)
[130] *Ibid*
[131] Lewis, P. Marsh, S. 'What is it like to trust a rock? A functionalist perspective on trust and trustworthiness in artificial intelligence'. (2022) Cognitive Systems Research. Vol 22. Pp.46.

at 8 p.m., But not trust the entity because the manufacturer may sell data to third parties or lack technical robustness. The AI Act favours the creation of AI systems people can trust especially for High-risk AI.

The trust problems outlined in the previous section are connected to widespread reservations about the impact of AI proliferation on society.[132] This has resulted in the establishment of a number of policy frameworks relating to Principled AI, frameworks to improve and regulate the Fairness, Accountability, and Transparency of AI-based services and products in particular.[133] Fjeld et al. examined existing Principled AI frameworks from industry, governments, the public, civic society, and academic institutions.[134] It is crucial to highlight that most ethical or principled AI frameworks focus on ethics, privacy, and other related issues rather than trust. Furthermore, the terms ethical machine learning and trustworthy machine learning are used interchangeably.[135]

There is the belief that trustworthiness is gained via adherence to ethical ideals such as human rights or non-discrimination policies. While ethical issues are integrally tied to trust judgments, ethical and trustworthy machine learning are not always synonymous. Ethical AI (machine learning) will unavoidably highlight the benevolence and integrity components of trust, while the other component of trust (ability and predictability) would go unnoticed.[136] Knowing what to do and doing it are two aspects of ethics. Trust is concerned with what or who should be trusted, as well as how to generate trust, irrespective of whether it is ethical.[137]

7. Conclusion

This chapter introduced the concepts that will form the basis of this study, establishing the ABI+ methodology for trustworthiness as the best way to define trust and trustworthiness. AI should be

[132] Greene D, Hoffmann AL and Stark L, 2019. "Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning" Proceedings of the Annual Hawaii International Conference on System Sciences

[133] Mittelstadt, B. "*Principles Alone Cannot Guarantee Ethical AI*" (2019).1 Nature Machine Intelligence 501

[134] Fjeld, J. Acten, N. Hilligoss, H. Naay, A. Slikumar, M. 'Principled AI: Mapping Consensus in Ethical and Rights – Based Approaches to Principles in AI'. (2020) Berkman Klein Center, Harvard University. Publication No 2020-1

[135] Koszegi, S.T. High-Level Expert Group on Artificial Intelligence. (2019)

[136] Toreini, E. Aitken, M. Coopamootoo, K. Elliott, K. Gonzalez Zelaya, C. and van Moorsel, A. 'The relationship between trust in AI and trustworthy machine learning technologies. (2020) In Conference on Fairness, Accountability, and Transparency (FAT* '20), Barcelona, Spain. ACM, New York, NY, USA, 12 pages. https://doi.org/10. 1145/n

[137] Raden N, "Trustworthy Ai versus Ethical AI - What's the Difference, and Why Does It Matter?" (*diginomica*May 11, 2021) <https://diginomica.com/trustworthy-ai-versus-ethical-ai-whats-difference-and-why-does-it-matter#:~:text=My%20take,whether%20or%20not%20it's%20ethical.> accessed August 3, 2022

capable, benevolent, trustworthy, and predictable. Ethical AI is marketed as trustworthy AI, but it is not because it emphasizes benevolence over trust, predictability, and integrity.[138] The HELG AI proposes that AI be lawful, robust, and ethical, further broken down to mean that AI systems must be fair, explainable, auditable, and safe, which the AI Act intends to achieve.

The goal of trustworthy AI regulation should be to ensure the development of AI systems that people can trust, as well as protection for them when things go wrong, to increase trust in the system.

[138] Lewis, P. Marsh, S. 'What is it like to trust a rock? A functionalist perspective on trust and trustworthiness in artificial intelligence'. (2022) Cognitive Systems Research. Vol 22. Pp.46.

Chapter III: VATs + Harms & Risks = Untrustworthiness

1. Untrustworthiness of VAT Systems

As discussed in the previous chapter, speech-based technology relies on allowing users to engage with devices and services using voice rather than keyboards and mouse movements. VAT services are enriched by the environment created by their suppliers, like as Amazon and Google. This ecosystem allows third-party developers to create new apps known as "Skills" on the Amazon ecosystem and "Actions" on the Google ecosystem to provide consumers with a more enriched experience.[139]

For the sake of this research, the third-party applications will be called "Skills". A skill is a system that processes the requests of a user by telling the VAT device what the response should be via a combination of a front-end interaction model and a backend cloud service code.[140].

A skill must be certified before it can be made publicly available. A Skill is certified when it has met the platform's policy guidelines, privacy, and security requirements. The certification process ensures that it adheres to the platform's content, privacy, and security regulations.[141] The process is a black box, because nobody can access the internal implementation process.[142]

The credibility of a skill certification is vital to platform providers (Amazon, Google, etc.), developers, and, most importantly, end-users.[143] Users have faith in VAT platforms to fulfil their

[139] Lentzsch, C. Shah, S.J. Andow, B. Degeling, M. Das, A. Enck, W. (2021) Hey Alexa, is this Skill Safe? Taking a Closer Look at the Alexa Skill Ecosystem. Network and Distributed Systems Security (NDSS) Symposium. https://www.ndss-symposium.org/wp-content/uploads/ndss2021_5A-1_23111_paper.pdf Accessed 23rd May 2022.
[140] Cheng, L. Wilson, C. Liao, S. Young, J. Dong, D. Hu, H. 'Dangerous Skills Got Certified: Measuring the Trustworthiness of Skill Certification in Voice Personal Assistant Platforms'. (2020) In Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security (CCS '20), November 9–13, 2020, Virtual Event, USA. ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3372297.3423339 Accessed 23rd May 2022.
[141] "Certification Requirements," Amazon Alexa, 2019. [Online]. Available: https://developer.amazon.com/en-US/docs/alexa/custom-skills/certification-requirements-for-custom-skills.html Accessed 6th June 2022.
[142] Cheng, L. Wilson, C. Liao, S. Young, J. Dong, D. Hu, H. 'Dangerous Skills Got Certified: Measuring the Trustworthiness of Skill Certification in Voice Personal Assistant Platforms'. (2020) In Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security (CCS '20), November 9–13, 2020, Virtual Event, USA. ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3372297.3423339 Accessed 23rd May 2022.
[143] Ibid

needs while protecting their privacy and security. Third-party developers believe that VAT platforms provide a trustworthy marketplace for them to promote their expertise and reach users.

Nevertheless, Cheng et al have found a weak vetting system in the certification process, leading to malicious skills getting certified and posing a significant threat to users of the VAT systems.[144] In their paper, Cheng et al. note that of the 234 (115 in the general category and 119 in the kids' category) skills they submitted for certification on the Amazon ecosystem, 55 of violated the content, privacy and security policy defined by Amazon were cleared and certified. On Google Ecosystem, of the 381 Actions submitted, 148 were certified, and 233 did not pass the certification process.[145] Cheng et al.'s research findings regarding these third-party platforms are that VAT manufacturers are more focused on quantity over the quality of the 3rd party apps. By prioritizing quantity over quality, gaps are bound to exist as there would be insufficient checks for the skills submitted for certification.[146]

These poor certification process of Skills create several risks that could affect trust. Research by Lentzsch et al. shows that the poor certification process of these Skills is the reason why these systems are prone to manipulations and attacks and breach of rights.[147] Zhang et al. showed that certified third-party apps are raising new types of security threats[148] – Voice squatting attack[149] and voice masquerading attack[150]. This means an attack on the system can be launched when two different Skills with similar pronunciation are called to the Alexa device. For example, where a

[144] Ibid
[145] Cheng, L. Wilson, C. Liao, S. Young, J. Dong, D. Hu, H. (2020) Dangerous Skills Got Certified: Measuring the Trustworthiness of Skill Certification in Voice Personal Assistant Platforms. In Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security (CCS '20), November 9–13, 2020, Virtual Event, USA. ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3372297.3423339 Accessed 23rd May 2022.
[146] Ibid
[147] Lentzsch, C. Shah, S.J. Andow, B. Degeling, M. Das, A. Enck, W. (2021) Hey Alexa, is this Skill Safe? Taking a Closer Look at the Alexa Skill Ecosystem. Network and Distributed Systems Security (NDSS) Symposium. https://www.ndss-symposium.org/wp-content/uploads/ndss2021_5A-1_23111_paper.pdf Accessed 23rd May 2022.
[148] Zhang, N. Mi, X. Feng, X. Wang, X. Tian, Y., and Qian, F. "Dangerous skills: Understanding and mitigating security risks of voice-controlled third-party functions on virtual personal assistant systems," in Proceedings of the 40th IEEE Symposium on Security and Privacy (SP), 2019, pp. 1381–1396.
[149] The Voice Squatting attack occurs when an adversary exploits how a skill is invoked (via voice command) and variations in how the command is spoken (e.g., phonetic differences caused by accent, courteous expression, etc.) to cause a VPA system to trigger a malicious skill rather than the one the user intends. Where a user says, *"Alexa, open Peppa Pig, please."* It should only open 'Peppa Pig' but can trigger a malicious skill called *'Peppa Pig Please'* once uploaded to the Skill market. Ibid
[150] The goal of the Voice Masquerading attack is to engage with the user and the VAT system. The system is meant to transfer all voice instructions to the currently operating skills, including those to be processed by the VAT system. To mimic the target skill and collect sensitive information from the user, a malicious skill might pretend to switch to another skill or terminate a running skill that is still running in response to the instruction.

user says, *"Alexa, open Peppa Pig, please."* It should only open 'Peppa Pig' but can trigger a malicious skill called *'Peppa Pig Please'* once uploaded to the Skill market. The device does not know which one the users want precisely, and the attacking skills can be installed on the device to collect data from users.[151]

Secondly, the privacy policy is a vital part of the submission paper for certification. Yet, there has been evidence to show inconsistencies between what the privacy policy states, and the data accessed.[152] For skills targeted at children and health and Fitness skills, only 13.6% of the skills in the kid's category have a privacy policy. Amazon does not mandate a privacy policy for skills targeted toward children under the age of 13.[153] 23% of the privacy policies do not fully disclose the type of data associated with permission requested by the Skill. Many skills that access full name permission (33%) did not disclose the collection of such in the privacy policy.[154]

Another discovery is that the certification allows skills targeted to kids have expletives in their contents. One of the Skills called "My Burn" said *"You are ugly you would scare the crap out of the toilet*," and the Skill called "New Fact" said *"A Pig orgasms last for 30 minutes".*[155] They found 33 skills for kids to have expletives in their content. They further found that 57.1% of parents lacked trust in the device because their children are exposed to expletives via the device.[156]

Relying on the ABI+ trust methodology, it is pertinent to note that the poor certification process that causes high security issues for users means that these systems cannot be trusted because they are lacking in specifically integrity and predictability.

2. **Harms raised by VAT and how trust may be violated**

[151] Zhang, N. Mi, X. Feng, X. Wang, X. Tian, Y., and Qian, F. "Dangerous skills: Understanding and mitigating security risks of voice-controlled third-party functions on virtual personal assistant systems," in Proceedings of the 40th IEEE Symposium on Security and Privacy (SP), 2019, pp. 1381–1396.
[152] Andow, B. Mahmud, S. Wang, Y W. Whitaker, J. Enck, W. Reaves, B. Singh, K., and Xie, T. "Policy Lint: Investigating Internal Privacy Policy Contradictions on Google Play," in Proceedings of the 28th USENIX Security Symposium (USENIX Security), 2019, pp. 585–602
[153] Zhang, N. Mi, X. Feng, X. Wang, X. Tian, Y., and Qian, F. "Dangerous skills: Understanding and mitigating security risks of voice-controlled third-party functions on virtual personal assistant systems," in Proceedings of the 40th IEEE Symposium on Security and Privacy (SP), 2019, pp. 1381–1396.
[154] *Ibid*
[155] Le, T., Huang, D.Y., Apthorpe, N.J., & Tian, Y. (2022). SkillBot: Identifying Risky Content for Children in Alexa Skills. *ACM Transactions on Internet Technology (TOIT)*. https://www.semanticscholar.org/reader/5895b85969544d455346f64f81ba62000907fbfd Accessed 23rd May 2022.
[156] Ibid

Where there are no adequate certification processes, there are risks connected with VAT usage, and a breach of confidence among system users.

Rogue VATs: Chung et al. demonstrate in their study that, while most communication with VAT systems is encrypted, not everything is communicated via a secure protocol, making it easy to discover a VAT device within a home.[157]

VAT must be linked to the internet to function. If they have security flaws, they are vulnerable to Distributed Denial of Service (DDoS) attacks, which occur when important online platforms and services become inaccessible to a significant number of users around the world.[158] A corrupted VAT can act as a virtual spy, which is a privacy risk. VATs are rapidly being connected to devices and appliances in the house like these via the internet of things, which implies that an assault on the system significantly harms the user's rights, wellness, and safety. The hacking of a VAT system of an individual by a disgruntled ex-partner is a prime illustration of such dangers.[159]

There have also been situations where the devices have recorded conversations in the home and sent them to random contact a thousand miles away.[160] There is also evidence to link the vulnerabilities with VAT systems and domestic abuse with partners having easy access to hack the system to hurt their partners.[161]

There is evidence that most VAT providers think that consumers would go to great lengths to protect their privacy, which is an unreasonable assumption.[162] Offloading the labour of guaranteeing system security on the user is a sort of victim blaming, and it reduces faith in these systems. For example, when Uber was hacked, they hid the fact from the public diminished the

[157] Chung H, Park J and Lee S, "Digital Forensic Approaches for Amazon Alexa Ecosystem" (2017) 22 Digital Investigation
[158] "DDoS Attack That Disrupted Internet Was Largest of Its Kind in History, Experts Say" (*The Guardian*October 26, 2016) <https://www.theguardian.com/technology/2016/oct/26/ddos-attack-dyn-mirai-botnet> accessed July 11, 2022
[159] Evans M, "Woman Accessed Ex-Partner's Alexa to Torment His New Girlfriend" (*The Telegraph*August 17, 2021) <https://www.telegraph.co.uk/news/2021/08/17/woman-accessed-ex-partners-alexa-device-tormented-new-partner/> accessed July 11, 2022
[160] "Little Did She Know, Alexa Was Recording Every Word She Said" (*NBCNews.com*May 24, 2018) <https://www.nbcnews.com/tech/tech-news/little-did-she-know-alexa-was-recording-every-word-she-n877286> accessed July 11, 2022
[161] "Terrifying Ways Partners Can Use Household Gadgets to Control and Spy on You" (*The Sun*March 16, 2022) <https://www.thesun.co.uk/news/17810998/tech-abuse-domestic-violence-refuge/> accessed July 11, 2022
[162] Aleisa N and Renaud K, "Privacy of the Internet of Things: A Systematic Literature Review" [2017] Proceedings of the 50th Hawaii International Conference on System Sciences (2017)

trust scale among users.[163]Kumar et al., for example, stressed that privacy and security threats have a considerable detrimental impact on perceived trust. They found out that a running belief among participants of their research is that these companies do not have the best interest of users, the easy infiltration by bad actors, and monitoring difficulty of these companies enhance distrust in VAT manufacturers.[164] To enhance trust, one must address the technology's risk perception.[165]

Using the ABI+ trust approach, the VAT system that is subject to attacks that prevent it from carrying out its tasks fails on both the ability, integrity, and reliability scales. For ability, a VAT subject to attack means that the device may not be able to perform the task the user has requested. For example, "Alexa! Turn off the light and the temperature", but the device cannot do so because a rogue skill is working to prevent the action in the background. The VAT lacks benevolence because the Skill's actions betray the user, for example, to monitor behaviour or harvest data. Failure in terms of integrity is that the Skill is created in bad faith. Such a VAT will become unreliable when it cannot perform the task assigned to it when needed.

Though the Act does not expressly define trustworthiness, it is a combination of attributes that indicate that an entity will not betray another due (benevolent) to bad faith (integrity) such as misaligned incentives, lack of care, disregard for promise keeping, or ineptitude at a task (ability and Reliability).[166] With the risks and harms VAT systems pose to users because of their flawed certification process, their trustworthiness is questioned.

3. **Conclusion**

The AI Act assures that trustworthy AI systems are available on the European market. There are vulnerabilities in the VAT ecosystem that hurt system users and hence fail to fulfil the standards that facilitate these systems to be trusted. It makes no sense to delegate responsibility for guaranteeing the reliability and safety of a VAT device to the user when the source of the vulnerabilities is the manufacturer. The shift of responsibility to users goes against the very concept of system trust and trustworthiness. Trust resides between the user and the provider; we

[163] Sarah Hospelhorn Based in Brooklyn and others, "Analyzing Company Reputation after a Data Breach" (*Varonis*) <https://www.varonis.com/blog/company-reputation-after-a-data-breach> accessed August 3, 2022
[164] Kumar D and others, "Emerging Threats in Internet of Things Voice Services" (2019) 17 IEEE Security & Privacy 18
[165] Ibid
[166] Levi, M. and Stoker, L, Political Trust and Trustworthiness' (2000) 3 Annual Review of Political Science 475

trust that the supplier supplied the product qualities, not that the product has the free will and choice to behave in a trustworthy manner.[167]

The next chapter will investigate the trustworthiness of VAT under the Act and the flaws in the Act that limit the development of trustworthy VAT systems.

[167] NíFhaoláin L, Hines A and Nallur V, "Assessing the Appetite for Trustworthiness and the Regulation of Artificial Intelligence in Europe" (*Assessing the Appetite for Trustworthiness and the Regulation of Artificial Intelligence in Europe | Research Repository UCD*January 7, 2021) <https://researchrepository.ucd.ie/handle/10197/12396> accessed August 3, 2022

Chapter IV: VATs and the AI ACT

"One should expect trust to be increasingly in demand as a means of enduring the complexity of the future which technology will generate."[168]

1. Trustworthiness of VAT and gaps in the AI Act.

Having established that there is a link between the poor certification process of skills, the risks and harms to users and the untrustworthiness of these systems, it is pertinent to look at how trustworthiness of VAT can be achieved via the AI Act.

The AI Act aims to protect individuals' safety and basic rights by utilizing a 'clearly defined' risk-based approach.[169] The European Commission gravitated toward a framework for high-risk AI systems, with the option for all suppliers of AI systems that are not high risk to adhere to a Code of Conduct (CoC).[170]

The Proposal exists to provide a method for routinely eliminating detrimental dangerous AI systems to keep AI trustworthy. This is demonstrated by the segmentation of AI system hazards inside the Proposal. Under the Act, AI is classified based on its risk:

a. Unacceptable AI: Any AI that poses a demonstrable risk to EU people shall be prohibited. Examples of such AI include social scoring by governments and toys that use voice assistance to urge youngsters to engage in risky behaviour. The proposed regulation prohibits some sorts of AI.
b. High Risk AI: These are AI systems utilized for important infrastructure, such as AI-powered transportation systems, which potentially endanger residents' lives. Examples, test scoring Product safety components, AI use in robot-assisted surgery, CV sorting software, AI systems for migration, asylum, and border control management, Justice administration and democratic procedures.

[168] Luhmann, N. 1979. Trust and power. Two works by Niklas Luhmann. Translated by Howard Davis. New York: John Wiley & sons Ltd.
[169] Stuurman, K. Lachaud, E. "Regulating AI: A Label to Complete the Proposed Act on Artificial Intelligence". (2022) Computer Law and Security Review 44, 105657. https://doi.org/10.1016/j.clsr.2022.105657.
[170] Custers B and Fosch-Villaronga E, *Law and Artificial Intelligence: Regulating AI and Applying AI in Legal Practice* (TMC Asser. Springer Science 2022)

c. Low Risks: These include AI systems such as chatbots or deepfakes. They are subject to only the most basic disclosure requirements. Technologies with low risks aim to provide users with the ability to make educated decisions about whether to continue or discontinue using the technology.

2. Transparency Obligation

VATs do not qualify as high-risk AI systems under this risk-based framework. However, they fall under Medium/Limited risk.[171] In the Act, medium-risk AI systems are expected to comply with a **Transparency Obligation**.[172] Article 52 provides that the transparency obligations apply to AI systems that (i) interact with humans, (ii) used to detect emotions or determine association with social categories based on the biometric data or generate or manipulate content (deep fakes).[173] Users must be told that they are engaging with an artificial intelligence system. This requirement does not apply if the interaction is obvious from the context of usage, according to the regulation. Due to the general design and marketing, the consumer should be aware that they are engaging with an AI system.

The problem with this exception is that VAT systems are becoming very sophisticated. For example, in 2011, Apple suggested to users to "talk to Siri as you would to a person" when Siri became a part of the iPhone operating system.[174] But VAT have metamorphosed into systems like the Google Duplex – that can book appointments on behalf of the user and because it interacts with third parties on behalf of the user there has to be strict adherence to the transparency obligation.

---

[171] Title IV of the Act

[172] Article 52, paragraph 4 of the Act.

[173] Article 52 provides when the transparency obligations apply:

1. Providers shall ensure that AI systems intended to interact with natural persons are designed and developed in such a way that natural persons are informed that they are interacting with an AI system, unless this is obvious from the circumstances and the context of use. This obligation shall not apply to AI systems authorized by law to detect, prevent, investigate, and prosecute criminal offences, unless those systems are available for the public to report a criminal offence.

2.Users of an emotion recognition system or a biometric categorization system shall inform of the operation of the system the natural persons exposed thereto. This obligation shall not apply to AI systems used for biometric categorization, which are permitted by law to detect, prevent and investigate criminal offences.

3.Users of an AI system that generates or manipulates image, audio or video content that appreciably resembles existing persons, objects, places or other entities or events and would falsely appear to a person to be authentic or truthful ('deep fake'), shall disclose that the content has been artificially generated or manipulated.

[174] MacArthur, E. (2014). The iPhone Erfahrung: Siri, the auditory unconscious, and Walter Benjamin's Aura. In D. M. Weiss, A. D. Propen, & C. Emmerson Reid (Eds.), Design, Mediation, and the Posthuman (pp. 113–127). Lanham: Lexington Books

On the other hand, the system providers (Amazon and Google) can argue that the context of the system does not require that the transparency obligation applies to their devices. As Smuha et al rightly stated, the Act is vague on the nature and manner of the information that should be submitted. The lack of defined criteria creates a significant danger since the information offered by system suppliers may differ.[175] This gap does not help build the trust of users of the system; if there are varying notifications or labels on VAT devices, consumers would be confused about what to look out for. Within the ABI+ the varying labels raises the question of predictability – differing labels with different meanings means becomes a predictability challenge. Furthermore, merely mandating AI providers to inform people that they are being subjected to intrusive technology does not address the chilling effects of these technologies rather it enhances them.[176] This indicates that the lack of standard information or labels for users may generate future issues since developers may employ language in diverse ways to comply with or bypass the legislation (lack of consistency), which does not aid in the trust-building process.

Critics have argued that this is compounded by the fact that the Act does not guarantee the public will receive sufficient information to understand these risks that they are being subjected to. There is also no clear pathway to contest the operation of certain AI systems and be able to use the information obtained to seek redress.[177]

3. Codes of Conduct

Apart from the transparency obligation, the other part of the Act that caters to VAT is the Codes of Conduct. A Code of Conduct is a policy that lays out the principles, standards, and the moral and ethical expectations of a company that employees and third parties are held to as they interact with the organization.[178] Article 69 distinguishes the types of codes of conduct for non-high-risk AI. The Code of Conduct, according to the Act, should either be drawn up by the European

[175] Stuurman, K. Lachaud, E. "Regulating AI: A Label to Complete the Proposed Act on Artificial Intelligence". (2022) Computer Law and Security Review 44, 105657. https://doi.org/10.1016/j.clsr.2022.105657.
[176] Smuha, N. Ahmed- Rengers, E. Harkens, A. Li, W. Maclaren, J. Pirelli, R. Yeung. (2021) How Can the EU Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for An Artificial Intelligence Act. LEADS Lab, University of Birmingham
[177] *Ibid*
[178] GAN Integrity. 2022. *What is Code of Conduct? | Definition | GAN Integrity*. [online] Available at: <https://www.ganintegrity.com/compliance-glossary/code-of-conduct/> [Accessed 13 June 2022].

Commission (the Commission) and a European Artificial Intelligence Board (the Board) or to encourage and facilitate the drawing up of such codes of conduct by system providers.[179]

This means that the Codes will apply to non-high-risk systems while being based on the standards established for high-risk systems in Title III, Chapter 2. (Requirements for High-Risk Systems). The drafters of the Codes must evaluate the technical requirements required to ensure compliance with Chapter 2 depending on the system's purpose.[180] The Act gives room for the systems providers - in this case VAT system providers - to draw up their code of conduct and, users and stakeholders can also be involved in the drafting process.[181]

4. *Problems with Codes of Conduct*

There are several issues with voluntary codes of conduct and how it contradicts the idea of fostering trustworthy AI systems within the Union. Regulation aims to provide a robust legal accountability mechanism.[182] Making voluntary codes of conduct does not provide a robust legal accountability mechanism for AI systems that do not qualify as High risk. Research on the impact of codes of conduct has yielded concerning results. Iacovina argues that codes are usually often followed in the letter than in spirit or as a checklist rather than as part of the indispensable reflexive practice of an organization.[183] According to research of the ACM Code of Ethics, it has minimal impact on the day-to-day decision making of software engineering professionals and students.[184] More research suggests that the existence of a code has no noticeable influence on unethical and dishonest behaviour.

The Act is supposed to be the legal accountability mechanism to foster the development of trustworthy AI systems within the Union. However, the Act is designed so that only systems that fall under the high-risk categorization benefit from legal accountability, leaving non-high-risk

[179] Article 69 of the Act-
[180] Article 69 paragraph 1 of the Act
[181] Article 69, paragraph 3 of the Act.
[182] Mittelstadt, B.C. "Principles Alone Cannot Guarantee Ethical AI". (2019). Nature Machine Intelligence. 1(11) https://www.researchgate.net/publication/337015694_Principles_alone_cannot_guarantee_ethical_AI/citations
[183] Iacovino, L. "Ethical Principles and Information Professionals: Theory, Practice and Education". (2002) Australian Academic & Research Libraries, 33:2, 57-74, DOI: 10.1080/00048623.2002.10755183
[184] McNamara, A., Smith, J. & Murphy-Hill, E. Does ACM's code of ethics change ethical decision making in software development? in Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering - ESEC/FSE 2018 729–733 (ACM Press, 2018). doi:10.1145/3236024.3264833

systems to be guided by codes of conduct that are not legally binding, other legislations like the Consumer Protection Rules[185] or federation of trade associations like the Direct Selling Europe[186].

Typically, the effect is only seen when the codes are ingrained in the company culture and aggressively enforced.[187] However, we have seen time and time again how easy it is for corporate entities to find ways to chuck out codes that stand in the way of the business model, for example, Google disbanding their ethical AI board[188] or firing staff who disagreed with development of unethical AI systems.[189] There is no guarantee or necessity that all, or even some, AI creators and users would follow the soft law suggestions.[190]

Penalties, particularly external sanctions for code violations, are critical to adherence and effective self-governance.[191] Since it is a voluntary code of conduct, there is no indication if there would be sanctions for non-compliance or breach.[192]

There is also the issue with harmonizing codes if system providers could develop their codes. The content of these codes, when considered side by side, will vary enormously due to different applications, and this is because organizations would usually modify the code not to fit the issues their systems raise but issues that are popular within news cycles.[193] Harmonization would be

[185] Directive (EU) 2019/2161 of the European Parliament and of the Council of 27 November 2019 amending Council Directive 93/13/EEC and Directives 98/6/EC, 2005/29/EC and 2011/83/EU of the European Parliament and of the Council as regards the better enforcement and modernisation of Union consumer protection rules
[186] "Code of Ethics" (*Direct Selling Europe*) <https://directsellingeurope.eu/about-us/code-ethics> accessed July 5, 2022
[187] Shilton, K. "That's Not an Architecture Problem!": Techniques and Challenges for Practicing Anticipatory Technology (2015) Ethics. 7.
[188] The Verge. 2022. *Google dissolves AI ethics board just one week after forming it*. [online] Available at: <https://www.theverge.com/2019/4/4/18296113/google-ai-ethics-board-ends-controversy-kay-coles-james-heritage-foundation> [Accessed 13 June 2022].
[189] The Verge. 2022. *Google dissolves AI ethics board just one week after forming it*. [online] Available at: <https://www.theverge.com/2019/4/4/18296113/google-ai-ethics-board-ends-controversy-kay-coles-james-heritage-foundation> [Accessed 13 June 2022].
[190] Marchant, G. "Soft Law: Governance of Artificial Intelligence". (2019). UCLA A.I. Pulse Papers. https://escholarship.org/content/qt0jq252ks/qt0jq252ks.pdf Accessed 28th May 2022.
[191] Filpovic, A. Koska,C & Paganini,C. "Developing a Professional Ethics for Algorithms: Learning from the Examples of Established Ethics". (2018) Bertelsmann Stifling. https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/Ethics_for_Algorithmists.pdf Accessed 25th May 2022.
[192] *Ibid*
[193] Pearson, C.E. Seyfang, G. "New Hope or Falso Dawn? Voluntary Codes of Conduct, Labour, Regulation and Social Policy in a Globalizing World" (2001). Global Social Policy 1(1) 48-78.

further complicated because of the wording of the codes. Codes usually contain broad generalizations and statements of intent with very few concrete and achievable standards.[194]

Subsequently, because of the complexity of the value chain in the VAT system – the provider, third-party app developers, third-party servers etc., it is not ascertainable to determine to who and to what extent are they covered by the Codes. This is illustrated by the VAT ecosystem certification process that VAT providers have in place to ensure third-party developers comply with regulatory and *'ethical'* standards and how they have failed to have a transparent and yielding process in the provision of trustworthy apps. The chances of a voluntary code of conduct existing to foster the development of trustworthy AI appears to be a farfetched idea. As Baker argued, if a company does not encourage its employees and partners to respect and adhere to the rules of conduct and efficiently monitor it, they remain futile.[195]

One of the unanticipated consequences of creating a code of conduct is increased public cynicism. It stems from the fact that reputational risk is one of the reasons why a VAT system provider would agree to adopt a code of conduct, assuming there is a label to indicate so for the benefit of users. Unfortunately, reputational hazards have weight for as long as the issue is in the eye of the public.[196] The fact that users must rely on fear of reputational harm or public protest to have their interests and rights treated seriously undermines the AI Act's aim, and this is made worse considering that the Act makes no provision for affected parties to seek redress.[197] Trustworthiness of these systems and the law comes into question specifically with regards to integrity and predictability, if users have to rely on public protests, no redress options for affected parties.

VAT are evolving quickly and may soon incorporate new characteristics that will have a significant influence on individuals' rights. Manufacturers of these systems can derive information from and about the users' environment, emotions, and behaviour, [198] and body conditions[199] based on the user's behaviour and tone of voice combined with third-party skills.

---

[194] *Ibid*
[195] Baker, M.B. "Promises and Platitudes: Toward a New 21st Century Paradigm for Corporate Codes of Conduct". (2007) Connecticut Journal of International Law, Volume 23, pp. 123-163.
[196] Parker, D.B. (1981) Ethical Conflicts in Computer Science and Technology.
[197] *Ibid*
[198] Crawford, K. (2021) Time to Regulate AI that Interprets Human Emotions. Blog Post entry Nature's World. https://www.nature.com/articles/d41586-021-00868-5 Accessed 24th May 2022.
[199] Fagherazzi, G. and others. (2021) Voice for Health: The Use of Vocal Biomarkers from Research to Clinical Practice. 5 Digital Biomarkers 78.

In this sense, they will also be exempt from the Act's necessary material requirements and limited to the transparency responsibility outlined in Article 52 and voluntary norms. This might change if the Commission uses its powers under Article 7 of the Act to enact a delegated act to amend the list of high-risk AI systems supplied in Annex III. Until then, just the information responsibilities will apply, and providers are encouraged to implement a code of conduct.[200]

5. Labelling and Verifications

According to Stuurman and Lachaud, establishing a verification method in addition to a label might be an alternative for addressing the challenges that occur with self-regulation based on codes of conduct. A voluntary European labelling scheme would highlight apps built on safe, responsible, and ethical AI and data, and therefore which applications to trust, allowing individuals affected to make an ethical decision.[201] In 2020, fourteen (14) EU member states pushed hard for the creation of a voluntary label that would "incentivize AI inventors and deployers to promote trustworthy AI proactively and systematically".[202]

Though the Labelling system has been in use in the EU for a while, it will pose many trust-related challenges for AI systems. For a variety of reasons, labels issued to VAT businesses that have self-declared to conform to a code of conduct or standard will confront trust challenges.[203] One of the trust challenges is that a label's content is intentionally limited because it serves as a shortcut to showing conformance with standards without discussing how the conformity was proved. This is analogous to skill certification, where the procedure is opaque. Furthermore, self-declaration of conformance does not ensure the candidate's real compliance, and the absence of enforcement commonly witnessed with self-regulation methods calls into question their trustworthiness. For example, the CJEU's judgment to invalidate the EU-US Privacy Shield has cast light on the ongoing lack of enforcement that plagues self-regulation schemes.[204] The EU used the Privacy

[200] Stuurman, K. Lachaud, E. (2022) Regulating AI: A Label to Complete the Proposed Act on Artificial Intelligence. Computer Law and Security Review 44, 105657. https://doi.org/10.1016/j.clsr.2022.105657.
[201] *Ibid*
[202] Non-Paper – Innovative and Trustworthy AI: Two sides of the same coin': Position Paper on behalf of Denmark, Belgium, the Czech Republic, Finland, France, Estonia, Ireland, Latvia, Luxembourg, the Netherlands, Poland, Portugal, Spain, and Sweden ON Innovative and Trustworthy AI. Available at https://www.permanentrepresentations.nl/documents/publications/2020/10/8/non-paper---innovative-and-trustworthy-ai Accessed 24th May 2022.
[203] Stuurman, K. Lachaud, E. (2022) Regulating AI: A Label to Complete the Proposed Act on Artificial Intelligence. Computer Law and Security Review 44, 105657. https://doi.org/10.1016/j.clsr.2022.105657
[204] Case C-311/18 Data Protection Commissioner v Facebook Ireland and Maximillian Schrems.

Shield to allow firms to send personal data from the EU to the US. The method requires US businesses to self-certify their adherence to the US Department of Commerce's list of data protection principles.

The lack of dependability affects trust, which is vital in label acceptance. The value of a label is determined by the level of trust in the issues and the scheme among end-users. If one party loses trust, it impacts the trust of the entire process and may jeopardize its sustainability.[205] Going by the ABI+ trust methodology, most especially on integrity and predictability. Users of the system may have difficulty trusting the process if the methodology used to evaluate/certify the systems is not reliable and integrable due to the existence of varying standards. Diverging methods mean a lack of coordination on what is the correct route to take, which may affect the trustworthiness of AI systems. Enforcement agencies withing Europe have a history of not having adequate resources for monitoring has left gaps in the labelling of goods and the differing national rules also affects the integrity of labels.

The use of CE markings has also proved that there is distrust in labels because some European manufacturers need to affix a national label alongside the CE label to reassure the public of the actual quality of their products.[206] This, in turn, may lead to a proliferation of labels which entails the risk of further confusing the public about the meaning of the labels. There is also the fact that most VAT companies are American companies and may also have to affix American labels. Once again on the ABI+ model, integrity and predictability are affected severely with the use of CE markings based on past experience with their application on other products.

Trust in the body that gives the label is also critical, raising concerns about the authenticity of the label's issuer. Having mentioned in an earlier chapter that there are various types of trust, Marchant et al. contend that reputational trust is the most common way for a consumer to gain sufficient security to engage in trusting behaviour, such as the purchase of an unfamiliar and unproven product simply because it is offered by a trusted company.[207] In its current form, the Act envisions an excessively large role for AI providers in the Regulation's implementation, providing them

[205] Stuurman, K. Lachaud, E. (2022) Regulating AI: A Label to Complete the Proposed Act on Artificial Intelligence. Computer Law and Security Review 44, 105657. https://doi.org/10.1016/j.clsr.2022.105657.
[206] Consumer Research Associates Ltd, Certification and Marks in Europe. (A Study Commissioned by the EFTA 2008) 16.
[207] Marchant, G.E. Sylvester, D.J and Abbott, K.W. (2010) A New Soft Law Approach to Nanotechnology Oversight: A Voluntary Product Certification Scheme. 28 UCLA Journal of Environmental Law and Policy 141.

excessive discretion and placing disproportionate reliance on conformity evaluations, codes of conduct, and CE marks. The key players in the VAT market are not exceptionally high up on the trust scale,[208] so this route does not seem like the best way to go about achieving trustworthy AI systems.[209] The trust scale here is the scale that the AI Act aims to achieve within the Regulation – that is the safeguarding of human rights and dignity.[210] Based on this, there is the challenge of achieving integrity, benevolence and predictability under the ABI+ model.

Finally, to establish legally trustworthy AI, the law must secure two key things: (i) the proper allocation and distribution of responsibility for AI-related wrongs and damages. (ii) a valid and efficient enforcement architecture that includes necessary transparency procedures to ensure effective protection of basic rights and the rule of law.[211]

The Act fails to recognise the status of individuals adversely affected by AI systems in its enforcement mechanisms, which is reflected in the total lack of procedural rights for individuals, such as a right to contest and seek redress, as well as a lack of adequate complaint mechanisms. The absence of rights for individuals reduces them to a passive entity, unaddressed and unacknowledged in the Act. This is very striking considering that one of the main reasons why AI is being regulated is to protect individuals from the risks of AI systems.[212]

Those who have their rights interfered with by the operation of a flawed VAT system are not granted the legal standing under the Act to initiate enforcement action for said interferences (for example, a VAT that hacked, and safety of user compromised) of its provisions, nor any enforceable legal rights for seeking mandatory order to bring an end to the violation or seek any form of remedy. Many individuals will be interested in seeking redress, and the absence of such an opportunity will cause further distrust in AI systems and regulations. It is possible to seek

[208] "Amazon Hit with $886M Fine for Alleged Data Law Breach" (*BBC News*July 30, 2021) <https://www.bbc.com/news/business-58024116> accessed August 2, 2022
[209] "Facebook and Google's Pervasive Surveillance of Billions of People Is a Systemic Threat to Human Rights" (*Amnesty International*August 17, 2021) <https://www.amnesty.org/en/latest/press-release/2019/11/google-facebook-surveillance-privacy/> accessed August 2, 2022
[210] Gentle J, "Amazon Surveillance Tech Customers Violate Human Rights" (*Open MIC*May 6, 2020) <https://www.openmic.org/news/relentlessly-reckless-amazon-surveillance-tech-customers-violate-human-rights> accessed August 2, 2022
[211] Stuurman, K. Lachaud, E. (2022) Regulating AI: A Label to Complete the Proposed Act on Artificial Intelligence. Computer Law and Security Review 44, 105657. https://doi.org/10.1016/j.clsr.2022.105657.
[212] Smuha, N. Ahmed- Rengers, E. Harkens, A. Li, W. Maclaren, J. Pirelli, R. Yeung. (2021) How Can the EU Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal For An Artificial Intelligence Act. LEADS Lab, University of Birmingham

remedy from existing regulations like the GDPR if it is a data related violation but relying on other existing regulation for a defective device defeats the idea of trustworthy devices because the root cause of the violation is not addressed.

The European Parliament recently produced research[213] that assessed whether European product liability law is sufficient to handle responsibility in AI. The author of that research concluded that the current approach is insufficient to handle the liability that arises in the context of AI.[214] Since many domains in which AI is and will be used are currently regulated independently at both the national and European levels.[215]

The Act's oversight, monitoring and enforcement regime fall short of the standard that legal trustworthiness requires. Users' trust and desire to acquire more innovative items increases with legal clarity and effective legal protection. Victims gain trust because they know they will always receive compensation if they are entitled to it. A victim is entitled to be reimbursed, at least partially, whenever she is not solely responsible for the damage sustained.[216] The result is that the Act is in danger of providing a façade of legal protection. It offers little meaningful and effective protection and has collapsed into a little more than self-regulation for systems that are non-high risk.

The Act fails to uphold the rule of law, and its enforcement is not on promulgated norms. This is because the current structure is complex and relies heavily on competent national authorities[217] which from experience shows they are not usually equally equipped with resources.[218] There would be the issue of unequal resources, and the four years post GDPR shows that monitoring and enforcement is significantly weakened due to lack of resources across the EU.[219]

---

[213] Bertolini, A.: Artificial Intelligence and Civil Liability. (2020), https://www.europarl. europa.eu/RegData/etudes/STUD/2020/621926/IPOL_STU(2020)621926_EN. Pdf

[214] NíFhaoláin L, Hines A and Nallur V, "Assessing the Appetite for Trustworthiness and the Regulation of Artificial Intelligence in Europe" (*Assessing the Appetite for Trustworthiness and the Regulation of Artificial Intelligence in Europe | Research Repository UCD*January 7, 2021) <https://researchrepository.ucd.ie/handle/10197/12396> accessed August 3, 2022

[215]Bertolini, A.: Artificial Intelligence and Civil Liability. (2020), https://www.europarl. europa.eu/RegData/etudes/STUD/2020/621926/IPOL_STU(2020)621926_EN. Pdf

[216] Ibid

[217] Article 59 of the Act.

[218] Masse, E. (2020) Two Years Under the EU GDPR: An Implementation Progress Report – State of Play, Analysis and Recommendations. AccessNow. https://www.accessnow.org/cms/assets/uploads/2020/05/Two-Years-Under-GDPR.pdf Accessed 13th June 2022.

[219] Ibid

6. Conclusion

This chapter has established why VATs are prone to difficulties that undermine users' rights through gaps in the certification process for third-party Skills, and how this affects trust in the AI system. It has gone so far as to establish that the current risk classification and regulatory mode for VAT under the Act ignore the concerns for adequate protection under EU law in the digital era, but, more importantly, it does not allow for the same standards to be applied to AI systems that do not qualify as High risk but pose a risk to users.

It is reasonable to argue that the Act, in its current form, does not inspire trust in VAT systems, and it leaves no provision for the protection and enforcement of the rights of persons impacted by these systems, thereby obliterating the purpose of developing trustworthy AI through legislation.

It is lacking specifically with regards to the application of ABI+ to VAT because the Act lacks the ability to provide meaningful and effective protection to users who are negatively impacted by the Act, and this trickles down to the fact that the absence of adequate enforcement mechanisms and human rights enforcements are no guaranteed brings the integrity of the Act into question.

Chapter V: Conclusion

A virtual assistant technology takes information and sophisticated data from conversations to comprehend and analyse them utilizing advanced Artificial Intelligence (AI), Robotic Process Automation (RPA), Natural Language Processing, and Machine Learning.[220]

VATs are known by different names describing the same technology – automated personal assistants, intelligent personal assistants, voice assistants, smart assistants, chatbots, intelligent automated assistants, virtual personal assistants, etc. There is also a growing market for conversational AI within and outside Europe, hence a need to consider their impact as an AI system, mainly because they cause harms that affect users, making them untrustworthy.

Relying on the ABI+ trust methodology, trustworthiness of VATs is impacted by ability, benevolence, integrity, and predictability of the system. The available research has found that manufacturers of VATs are not taking the necessary steps to ensure that users' safety, privacy, and security are protected—the risks and harms raised by VAT systems, and this impacts the trustworthiness of the systems.

Within the ABI+ trust framework, VATs fail to determine the intentions of a corrupted Skill (Ability and Integrity). In some cases, the Skill – (a system that processes the requests of a user by telling the VAT device what the response should be via a combination of a front-end interaction model and a backend cloud service code) may be able to perform users' instructions but lacks benevolence, integrity, and predictability. While in other situations, it lacks all four aspects of trustworthiness.

An example is that a VAT device with a rogue Skill may still be able to perform every task requested by the user, but the skill is present to harvest data or monitor behaviour of the user. Such a device lacks in integrity, benevolence, and predictability. On the other hand, another Skill can be on the device that makes it impossible for the device to perform task requested by the user, then such a device fails on the four scales of trust.

[220] Person, “Top 10 AI-Powered Virtual Assistant Companies” (*AI Magazine*April 26, 2022) <https://aimagazine.com/ai-applications/top-10-ai-powered-virtual-assistant-companies> accessed July 12, 2022

Trustworthy AI can uphold a task without compromising the user. Several well-known VAT systems are not trustworthy with the lax certification processes involved in the VAT ecosystems and the risks that occur because of these lax certification processes.

Trustworthy AI according to HLEG- has three (3) components that must be met throughout the lifecycle of an AI system. An AI system must be lawful – complying with all applicable laws and regulations. It should be technically and socially ethical, stick to ethical principles and values, and be robust. AI systems can cause unintentional harm even with good intentions.[221]

Based on the categorization of risks within the AI Act, VAT fall under medium/low risk, and the compliance requirement for systems that fall within this category is the transparency obligation that requires providers of the system to inform the user that they are interacting with an AI system. The Act also proposes a voluntary Code of Conduct that incorporates the requirements that High-risk AI systems are expected to comply with. The code of conduct having the exact requirements for high-risk systems is a good idea. However, the challenge is that the codes are voluntary – meaning a manufacturer can decide not to apply them and acknowledge that codes of conduct do not impact regulators think they have.

The Draft AI Act contains an uncommon mismatch between its substantive obligations' objective (mainly high-risk systems) and its material reach (all AI systems). The Draft AI Act, on the other hand, intends to both develop uniform standards and to exclude a wide range of software from further constraints without imposing any of its own.[222]

The Act does not make any provisions for users who have been negatively affected by VAT systems to seek redress and cannot initiate enforcement action for violations of the Act. This defeats the purpose of having a regulation that upholds the rights of citizens of the Union. This calls into question the relevance of 'trustworthy' designed regulation if users are not protected from the impact of these technologies.

[221] European Commission. (2019) High-Level Expert Group on Artificial Intelligence. https://ec.europa.eu/digital-single-maret/en/high-level-expert-group-artificial-intelligence Accessed 13th April 2022.
[222] Veale, M and Borgesius, FZ. 'Demystifying the Draft EU Artificial Intelligence Act' [2021] 22(4) Computer Law Review International. https://arxiv.org/abs/2107.03721v2 Accessed 27th September 2021

The draft Act may create a wide divide between high risks AI systems which are regulated, and non-high-risk systems which member states are blocked form regulating. Furthermore, the Act does very little to reduce the fundamental rights risks especially of systems not covered under Annex III. It is difficult to find the logic between the rules for some AI systems and then non-existent/weak rules for other types of AI that can also pose risks to users.

The limitations of the Act go further by not taking into consideration the rights of individuals affected by AI systems. As only those with obligations (manufacturers) under the Draft AI Act can challenge regulators' decisions, rather than those whose fundamental rights deployed AI systems affect, the Draft AI Act lacks a bottom-up force to hold regulators to account for weak enforcement. The GDPR which allows impacted organizations to file complaints, is already marked by slowness and inertia. As a result, enforcement of the Draft AI Act appears to be much less promising than it has been with the GDPR thus far.[223]

Furthermore, the incoherence of the enforcement system put in place in the Act places a lot of responsibilities on the member states to monitor, investigate and research obligations for AI that fall under the transparency obligations, and this is different from product regulation. Also, there is no guarantee that the supervisory authority of member states is going to be independent like it is required under the GDPR.

The Act does state that user requirements for high-risk systems are "without prejudice to other user obligations under Union or national legislation."[224] However, no analogous clause exists that applies to the whole ambit of the Draft AI Act, which is itself vast. As a result, there is legal confusion over whether current national algorithmic transparency standards that reach beyond 'high-risk' systems, such as those in the French public sector, would have to be disapplied.[225]In conclusion, the Act is a welcome and forward-thinking regulation. However, the gaps within it make it difficult to see how AI systems' trustworthiness, especially of VATs, can be achieved via

[223] See European Parliament resolution of 25 March 2021 on the Commission evaluation report on the implementation of the General Data Protection Regulation two years after its application (2020/2717(RSP)) para 17 (on regulatory paralysis in data protection enforcement)
[224] Art 29 (2) AI Act,
[225] Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique; décret n° 2017-330 du 14 mars 2017 relatif aux droits des personnes faisant l'objet de décisions individuelles prises sur le fondement d'un traitement algorithmique, art 1. See further Lilian Edwards and Michael Veale, 'Enslaving the Algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"?' (2018) 16 IEEE Security & Privacy 46, 48.

the Act since the requirements for VAT within the Act are a mere transparency obligation, voluntary codes of conduct, and labelling.

The idea of not relying on just ethics to solve AI (VAT) challenges and establishing a framework for trustworthy AI that combines ethics and law – trustworthy AI should be ethical, lawful, and robust is a solid foundation. However, the current structure of the Act may yield the same constraints that other options like mere ethics have yielded and the goal of trustworthy may not be achieved.

The evidence from the research shows that these are not viable mechanisms to ensure systems are built to protect the users. So, in answering the research question – it seems very unlikely that trustworthiness of VAT systems maybe achieved via the Act.

BIBLIOGRAPHY

Laws and Regulations

Directive (EU) 2019/2161 of the European Parliament and of the Council of 27 November 2019 amending Council Directive 93/13/EEC and Directives 98/6/EC, 2005/29/EC and 2011/83/EU of the European Parliament and of the Council as regards the better enforcement and modernisation of Union consumer protection rules

European Commission, Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM (2021) 206 final) (hereafter 'AI Act').

Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on Medical Devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and the repealing Council Directives 90/385/EEC and 93/42/EEC (OJ L 117, 5.5.2017, 1.

Case Law

Case C-311/18 Data Protection Commissioner v Facebook Ireland and Maximillian Schrems.

European Commission Publications/Guidelines

"Intelligent Automation Assistant - Patent HK-1220023-A1 - Pubchem" (*National Center for Biotechnology Information. PubChem Compound Database*) <https://pubchem.ncbi.nlm.nih.gov/patent/HK-1220023-A1> accessed June 16, 2022.

Consumer Research Associates Ltd, Certification and Marks in Europe. (A Study Commissioned by the EFTA 2008) 16.

China Academy of Information and Communications Technology (CAICT; 中国信息通信研究院; 中国信通院) and JD Explore Academy, "White Paper on Trustworthy Artificial Intelligence" (2021)

European Commission High Level Expert Group on Artificial Intelligence, Ethics Guidelines for Trustworthy AI, European Commission 2019. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai Accessed 26th September 2021

European Commission, 'The Rise of Virtual Personal Assistants' 2018, Digital Transformation Monitor, https://ati.ec.europa.eu/sites/default/files/202005/The%20rise%20of%20Virtual%20Personal%20Assistants%20%28v1%29.pdf Accessed on the 18th September 2021.

European Commission. Europe fit for the Digital Age: Commission proposes new rules and actions for excellence and trust in Artificial Intelligence. 2021. https://ec.europa.eu/commission/presscorner/detail/en/IP_21_1682 Accessed 23rd February 2022.

Leslie, D. (2019). Understanding Artificial Intelligence Ethics and Safety: A Guide for the Responsible Design and Implementation of AI Systems in the Public Sector. The Alan Turing Institute. https://www.turing.ac.uk/sites/default/files/2019-08/understanding_artificial_intelligence_ethics_and_safety.pdf

Masse, E. (2020) Two Years Under the EU GDPR: An Implementation Progress Report – State of Play, Analysis and Recommendations. AccessNow. https://www.accessnow.org/cms/assets/uploads/2020/05/Two-Years-Under-GDPR.pdf Accessed 13th June 2022.

Non-Paper – Innovative and Trustworthy AI: Two sides of the same coin': Position Paper on behalf of Denmark, Belgium, the Czech Republic, Finland, France, Estonia, Ireland, Latvia, Luxembourg, the Netherlands, Poland, Portugal, Spain, and Sweden ON Innovative and Trustworthy AI. Available at https://www.permanentrepresentations.nl/documents/publications/2020/10/8/non-paper---innovative-and-trustworthy-ai Accessed 24th May 2022.

United Nations Educational, Scientific, and Cultural Organization (UNESCO) and the World Commission on the Ethics of Scientific Knowledge and Technology. 2019. Preliminary Study on the Ethics of Artificial Intelligence. https://unesdoc.unesco.org/ark:/48223/pf0000367823 Accessed 26th January 2022.

Academic Literature

Aleisa N and Renaud K, "Privacy of the Internet of Things: A Systematic Literature Review" [2017] Proceedings of the 50th Hawaii International Conference on System Sciences (2017)

Andow, B. Mahmud, S. Wang, Y W. Whitaker, J. Enck, W. Reaves, B. Singh, K., and Xie, T. "Policy Lint: Investigating Internal Privacy Policy Contradictions on Google Play," in Proceedings of the 28th USENIX Security Symposium (USENIX Security), 2019, pp. 585–602

Ashraf C, "Exploring the Impacts of Artificial Intelligence on Freedom of Religion or Belief Online" (2021) 26 The International Journal of Human Rights 757

Baier A, "Trust and Antitrust" (1986). 96 Ethics 231

Baker, M.B. "Promises and Platitudes: Toward a New 21st Century Paradigm for Corporate Codes of Conduct". (2007) Connecticut Journal of International Law, Volume 23, pp. 123-163.

Bauer PC, "Clearing the Jungle: Conceptualizing and Measuring Trust and Trustworthiness" (2013). SSRN Electronic Journal

Bauer, P.C. "Conceptualizing Trust and Trust Worthiness", (2019) Research Gate. https://www.researchgate.net/publication/262258778_Conceptualizing_Trust_and_Trustworthiness

Bolton T and others, "On the Security and Privacy Challenges of Virtual Assistants" (2021) 21 Sensors 2312

Cheng, L. Wilson, C. Liao, S. Young, J. Dong, D and Hu. H. "Dangerous Skills Got Certified: Measuring the Trustworthiness of Skill Certification in Voice Personal Assistant Platforms". (2020). Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security. Association for Computing Machinery, New York, NY, USA, 1699–1716. https://doi.org/10.1145/3372297.3423339 Accessed 26th May 2022

Chung H, Park J and Lee S, "Digital Forensic Approaches for Amazon Alexa Ecosystem" (2017) 22 Digital Investigation

Dietz G and Den Hartog DN, "Measuring Trust inside Organisations" (2006) 35 Personnel Review 557

Ellul, J. Should we regulate Artificial Intelligence or some uses of software? *Discov Artif Intell* **2,** 5 (2022). https://doi.org/10.1007/s44163-022-00021-9

Epstein, J; Klinkenberg, W. D, "From Eliza to Internet: A Brief History of Computerized Assessment" (2001) *Computers in Human Behavior*. **17** (3): 295–314. doi:10.1016/S0747-5632(01)00004-8. ISSN 0747-5632.

Fagherazzi, G. and others. "Voice for Health: The Use of Vocal Biomarkers from Research to Clinical Practice". (2021) 5 Digital Biomarkers 78.

Filpovic, A. Koska,C & Paganini,C. "Developing a Professional Ethics for Algorithms: Learning from the Examples of Established Ethics". (2018) Bertelsmann Stifling. https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/Ethics_for_Algorithmists.pdf Accessed 25th May 2022.

Fjeld, J. Acten, N. Hilligoss, H. Naay, A. Slikumar, M.. "Principled AI: Mapping Consensus in Ethical and Rights – Based Approaches to Principles in AI". (2020) Berkman Klein Center, Harvard University. Publication No 2020-1

Floridi, L. "Establishing the rules for building trustworthy AI". (2019) *Nature Machine Intelligence, 1*(6), 261–262. https://doi.org/10.1038/s42256-019-0055-y.

Følstad A, Nordheim CB and Bjørkli CA, "What Makes Users Trust a Chatbot for Customer Service? an Exploratory Interview Study" [2018] Internet Science 194

Gal, M.S. "Algorithmic Challenges to Autonomous Choice", (2018) Michigan Telecoms and Tech Law Review.

Geiger C and others, "Testable Design Representations for Mobile Augmented Reality Authoring" Proceedings. International Symposium on Mixed and Augmented Reality, pp. 145-146.

Goskel-Canbek, N. and Mutlu, M.E. "On the Track of Artificial Intelligence: Learning with Intelligent Personal Assistants". (2016). *International Journal of Human Sciences*, 13 (1), 592 - 601.

Greene D, Hoffmann AL and Stark L, "Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning" (2019). Proceedings of the Annual Hawaii International Conference on System Sciences

Hancock, P.A., Kessler, T.T., Kaplan, A.D., Brill, J.C., & Szalma, J.L. "Evolving Trust in Robots: Specification Through Sequential and Comparative Meta-Analyses". (2020) *Human Factors*, pp. 18720820922080-18720820922080. https://doi.org/10.1177/0018720820922080

Iacovino, L. "Ethical Principles and Information Professionals: Theory, Practice and Education". (2002) Australian Academic & Research Libraries, 33:2, 57-74, DOI: 10.1080/00048623.2002.10755183

In Conflict, Cooperation, and Justice: Essays Inspired by the Work of Morton Deutsch, edited by Barbara B Bunker and Jeffrey Z Rubin. San Francisco, CA: Jossey-Bass. Bromiley, P., Cummings, L.L. (1995). "Transactions Costs in Organizations with Trust." Research on Negotiation in Organizations 5: 219–50.

Jacovi A and others, "Formalizing Trust in Artificial Intelligence" [2021] Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency https://doi.org/10.1145/3442188.3445923 .

Kastner, L. On Relation of Trust and Explainability: Why Engineer for Trust. https://arxiv.org/pdf/2108.05379.pdf

Keymolen E and Van der Hof S, (2019. "Can I Still Trust You, My Dear Doll? A Philosophical and Legal Exploration of Smart Toys and Trust" 4 Journal of Cyber Policy 143

Keymolen E, "Trust in the Networked Era" (2018) 22 Techné: Research in Philosophy and Technology 51

Korot E., Wagner S.K., Faes L., Liu X., Huemer J., Ferraz D., Keane P.A., Balaskas K. "Will AI replace ophthalmologists?" (2020) Translational Vision Science & Technology, 9 (2)

Koszegi, S.T. "High-Level Expert Group on Artificial Intelligence". (2019).

Kumar D and others, "Emerging Threats in Internet of Things Voice Services" (2019) 17 IEEE Security & Privacy 18

Lamontagne, L. Laviolette, F. Khoury, R. Bergen-Guyard, A. "A Framework for Building Adaptive Intelligent Virtual Agents", (2014). 15h IASTED International Conference on AI and Applications, pp. 17-19

Le, T., Huang, D.Y., Apthorpe, N.J., & Tian, Y. “SkillBot: Identifying Risky Content for Children in Alexa Skills”. (2022) ACM *Transactions on Internet Technology (TOIT)*. https://www.semanticscholar.org/reader/5895b85969544d455346f64f81ba62000907fbfd Accessed 23rd May 2022.

Lentzsch, C. Shah, S.J. Andow, B. Degeling, M. Das, A. Enck, W. “Hey Alexa, is this Skill Safe? Taking a Closer Look at the Alexa Skill Ecosystem”. (2021) Network and Distributed Systems Security (NDSS) Symposium. https://www.ndss-symposium.org/wp-content/uploads/ndss2021_5A-1_23111_paper.pdf Accessed 23rd May 2022.

Levi, M. and Stoker, L, Political Trust and Trustworthiness’ (2000) 3 Annual Review of Political Science 475

Lewicki RJ and Bunker BB, “Developing and Maintaining Trust in Work Relationships” [1996] Trust in Organizations: Frontiers of Theory and Research 114.”

Lewis, P.R. Marsh, S. ‘What is it like to trust a rock: A functionalist perspective on trust and trustworthiness in AI?’ (2022) Cognitive Systems Research. Vol 72, pg.33-49.

Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., Yi, J., & Zhou, B. “Trustworthy AI: From Principles to Practices” (2021) *ArXiv, abs/2110.01167*.

Lopes, G. Quesada, L. Guerrero, L.A. “Alexa vs Siri, Cortana vs Google Assistant: A Comparison of Speech Based Natural User Interface”, (2018) International Conference on Applied Human Factors and Ergonomics, Springer, 241-250.

Marchant, G. “Soft Law: Governance of Artificial Intelligence”. (2019). UCLA A.I. Pulse Papers. https://escholarship.org/content/qt0jq252ks/qt0jq252ks.pdf Accessed 28th May 2022.

Marchant, G.E. Sylvester, D.J and Abbott, K.W. “A New Soft Law Approach to Nanotechnology Oversight: A Voluntary Product Certification Scheme”. (2010) 28 UCLA Journal of Environmental Law and Policy 141.

Mayer RC, James DH, and David SF, “An Integrative Model of Organizational Trust - JSTOR” (July 1995) <https://www.jstor.org/stable/258792> accessed June 16, 2022, 33. Pp. 225-232

McCarthy, P. 'Mathematical logic in artificial intelligence' [1988] 117(1) Daedalus, Journal of the American Academy of Arts and Sciences 297-311

McNamara, A., Smith, J. & Murphy-Hill, E. “Does ACM’s code of ethics change ethical decision making in software development?” (2018) in Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering - ESEC/FSE 2018 729–733 (ACM Press, 2018). doi:10.1145/3236024.3264833

Mishra, A.K. “Organizational responses to crisis: the centrality of trust”, (1996), in Kramer, R.M. and Tyler, T.R. (Eds), Trust in Organisations: Frontiers of Theory and Research, Sage, Thousand Oaks, CA, pp. 261-87

Mittelstadt, B. 2019. “Principles Alone Cannot Guarantee Ethical AI” .1 Nature Machine Intelligence 501

Mutrak, A. Patil, S. Ticlke, A. Nimbalkar, A. Yadav, S. “Intelligent Virtual Assistant – Vision”. (2021). International Journal for Research in Applied Science & Engineering Technology, 2021

N. K and others, “Intelligent Personal Assistant - Implementing Voice Commands Enabling Speech Recognition” [2020] 2020 International Conference on System, Computation, Automation and Networking (ICSCAN)

Obermeyer, Z., Powers, B., Vogeli, C. and Mullainathan, S., “Dissecting racial bias in an algorithm used to manage the health of populations”. (2019) *Science*, 366(6464), pp.447-453.

Pandey, A. Vashist, V. Tiwari, P. Sikka, S. Makkar, P. “Smart Voice Based Virtual Personal Assistants with Artificial Intelligence”. (2020) Artificial and Computational Intelligence, Vol 1, Issue 3.

Parker, D.B. “Ethical Conflicts in Computer Science and Technology”. (1981)

Pearson, C.E. Seyfang, G. “New Hope or Falso Dawn? Voluntary Codes of Conduct, Labour, Regulation and Social Policy in a Globalizing World” (2001). Global Social Policy 1(1) 48-78.

Prabhu, V., Taaffe, K., & Pirrallo, R.. “Multi-Layered LSTM for Predicting Physician Stress During an ED Shift”. (2020) IIE Annual Conference. Proceedings, 1223.

Seppanen, R. Bloomqvist, K. Sundquvist, S. “Measuring Inter-Organizational Trust – A Critical Review of the Empirical Research in 1990-2003”. (2007) Industrial Marketing Management 36 (2) 249 -258.

Shilton, K. “That’s Not an Architecture Problem!”: Techniques and Challenges for Practicing Anticipatory Technology (2015) Ethics. 7.

Siau, K., & Wang, W. “Building Trust in Artificial Intelligence”, (2018) Machine Learning, and Robotics. *Cutter Business Technology Journal*, 31(2), pp. 47-53.

Silva, A and others, 'Intelligent Personal Assistants: A Systematic Literature Review' [2020] 147(1) Expert Systems with Applications, https://doi.org/10.1016/j.eswa.2020.113193 Accessed 26th September 2021

Smuha, N.A & Ahmed-Rengers, E & Harken, A & Li, W & MacLaren, J, & Piselli, R & Yeung, K. (2021). How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission’s Proposal for An Artificial Intelligence Act. LEADS Lab, Birmingham University. Accessed 28th April 2022.

Smuha, N.A. 'The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence' (2019) 20(4) Computer Law Review International 97-106. https://doi.org/10.9785/cri-2019-200402 Accessed 28th September 2021

Stucke ME and Ezrachi A, “How Your Digital Helper May Undermine Your Welfare, and Our Democracy” [2017] SSRN Electronic Journal.

Stuurman, K. Lachaud, E. (2022) Regulating AI: A Label to Complete the Proposed Act on Artificial Intelligence. Computer Law and Security Review 44, 105657. https://doi.org/10.1016/j.clsr.2022.105657.

Terzopoulos, G. and Satratzemi, M. “Voice Assistants and Smart Speakers in Everyday Life and in Education”. (2020). Informatics in Education. Vol 19, No 3, 473-490.

Thiebes, S., Lins, S. & Sunyaev, A. “Trustworthy artificial intelligence”. (2021). *Electron Markets* **31,** 447–464 https://doi.org/10.1007/s12525-020-00441-4

Toreini, E. Aitken, M. Coopamootoo, K. Elliott, K. Gonzalez Zelaya, C. and van Moorsel, A. "The relationship between trust in AI and trustworthy machine learning technologies". (2020) In Conference on Fairness, Accountability, and Transparency (FAT* '20), Barcelona, Spain. ACM, New York, NY, USA, 12 pages. https://doi.org/10. 1145/n

Tschopp PQ-Rand M, "Relationship between Trust and Law Is Counterintuitive and Paradox" (Can Laws build Trust in AI?), https://www.scip.ch/en/?labs.20210916&gt ; accessed December 12, 2021

Tur G, Deoras A and Hakkani-Tür D, "Detecting out-of-Domain Utterances Addressed to a Virtual Personal Assistant" [2014] Interspeech 2014 pp. 283 -287. pp. 283 -287.

Veale M and Zuiderveen Borgesius F, "Demystifying the Draft EU Artificial Intelligence Act — Analysing the Good, the Bad, and the Unclear Elements of the Proposed Approach" (2021) 22 Computer Law Review International 97

Veale, M and Borgesius, FZ. 'Demystifying the Draft EU Artificial Intelligence Act' [2021] 22(4) Computer Law Review International. https://arxiv.org/abs/2107.03721v2 Accessed 27th September 2021

Verspoor, K. and Cohen, K., *Natural Language Processing.* (2013). *Encyclopedia of Systems Biology, pp.1495-1498.*

Wang, P. 'On Defining Artificial Intelligence'. (2019) Journal of Artificial General Intelligence 10(2). Pp. 1-37

Yaghoubzadeh, K. Kramer, M. Pitsch, K. Kopp, S. "Virtual Agents as Daily Assistants for Elderly or Cognitively Impaired People", (2013), International Workshop on Intelligent Virtual Assistants, Springer, Berlin. PP. 79-91.

Zhang, N. Mi, X. Feng, X. Wang, X. Tian, Y., and Qian, F. "Dangerous skills: Understanding and mitigating security risks of voice-controlled third-party functions on virtual personal assistant systems," in Proceedings of the 40th IEEE Symposium on Security and Privacy (SP), 2019, pp. 1381–1396.

Books

Custers B and Fosch-Villaronga E, *Law and Artificial Intelligence: Regulating AI and Applying AI in Legal Practice* (TMC Asser 2022)

Dingli A, Haddod F and Klüver Christina, *Artificial Intelligence in Industry 4.0: A Collection of Innovative Research Case-Studies That Are Reworking the Way We Look at Industry 4.0 Thanks to Artificial Intelligence* (Springer 2022)

Frankish, K. and Ramsey, W., Cambridge Handbook of Artificial Intelligence (1st edn, Cambridge University Press 2014) p.15

Freitas, R. and Iacono, S., 2021. *Trust Matters*. London: Bloomsbury Publishing Plc.

Hardin R, *Trust and Trustworthiness* (Russell Sage Foundation 2002)

Hauk N and Hauswirth M, 1970. "Navigating through Changes of a Digital World" (*SpringerLink* January 1,) <https://link.springer.com/chapter/10.1007/978-3-030-86144-5_37> accessed June 20, 2022

Li B and others, *Advanced Data Mining and Applications 17th International Conference, ADMA 2021, Sydney, NSW, Australia, February 2-4, 2022, Proceedings, Part I* (Springer International Publishing 2022)

Luhmann, N. 1979. "Defining the Problem: Social Complexity," in Trust and Power, John Wiley & Sons, pp. 5 - 11.

Luhmann, N. 1979. Trust and power. Two works by Niklas Luhmann. Translated by Howard Davis. New York: John Wiley & sons Ltd.

MacArthur, E. (2014). The iPhone Erfahrung: Siri, the auditory unconscious, and Walter Benjamin's Aura. In D. M. Weiss, A. D. Propen, & C. Emmerson Reid (Eds.), Design, Mediation, and the Posthuman (pp. 113–127). Lanham: Lexington Books

Mayer-Schönberger, Viktor and Cukier K, *Big Data: A Revolution That Will Transform How We Live, Work and Think* (John Murray 2017)

Möllering Guido, *Trust: Reason, Routine, Reflexivity* (Emerald 2008)

Nilsson NJ, *The Quest for Artificial Intelligence: A History of Ideas and Achievements* (Cambridge University Press 2010

Pieraccini, R. (2017) AI Assistants, the MIT Press Essential Knowledge Series, MIT Press. Pp. 22

Russell SJ and Norvig P, *Artificial Intelligence: A Modern Approach* (Pearson 2022)

Stahl, B.C. (2021). Artificial Intelligence for a Better Future. Springer Briefs in Research and Innovation Governance, Chapter 4. Pp. 35

Turner J, *Robot Rules: Regulating Artificial Intelligence* (Palgrave Macmillan 2019)

Weizenbaum, J. (1976). *Computer power and human reason: from judgment to calculation*. Oliver Wendell Holmes Library Phillips Academy. San Francisco: W. H. Freeman

Articles and Websites

(*Advanced Technologies for Industry*) <https://ati.ec.europa.eu/> accessed June 16, 2022

(Directed by Google Developers YouTube 2018); https://www.youtube.com/watch?v=ogfYd705cRs&gt ; 35.04 -40:15, Accessed December 12, 2021

"Automated Personal Assistant" (*Wikipedia*July 20, 2021) <https://en.wikipedia.org/wiki/Automated_personal_assistant#:~:text=An%20automated%20personal%20assistant%20or,sources%20(such%20as%20weather%20conditions%2C> accessed June 16, 2022

"Certification Requirements," Amazon Alexa, 2019. [Online]. Available: https://developer.amazon.com/en-US/docs/alexa/custom-skills/certification-requirements-for-custom-skills.html Accessed 6th June 2022.

"Code of Ethics" (*Direct Selling Europe*) <https://directsellingeurope.eu/about-us/code-ethics> accessed July 5, 2022

"DDoS Attack That Disrupted Internet Was Largest of Its Kind in History, Experts Say" (*The Guardian*October 26, 2016) <https://www.theguardian.com/technology/2016/oct/26/ddos-attack-dyn-mirai-botnet> accessed July 11, 2022

"Little Did She Know, Alexa Was Recording Every Word She Said" (*NBCNews.com*May 24, 2018) <https://www.nbcnews.com/tech/tech-news/little-did-she-know-alexa-was-recording-every-word-she-n877286> accessed July 11, 2022

"Terrifying Ways Partners Can Use Household Gadgets to Control and Spy on You" (*The Sun*March 16, 2022) <https://www.thesun.co.uk/news/17810998/tech-abuse-domestic-violence-refuge/> accessed July 11, 2022

"The History of Chatbots - from Eliza to Alexa" (*AI Chatbot Platform from Onlim*, December 3, 2021) https://onlim.com/en/the-history-of-chatbots/ . Accessed December 12, 2021.

AI Multiple. 2022. [online] Available at: <https://research.aimultiple.com/conversational-ui/> [Accessed 15 June 2022].

BBC News. (2021). Alexa tells 10-year-old girl to touch live plug with penny. https://www.bbc.com/news/technology-59810383 Accessed 5th February 2022.

BBC News. 2022. *High-frequency trading and the $440m mistake*. [online] Available at: <https://www.bbc.com/news/magazine-19214294> [Accessed 15 June 2022].

Bloomberg.com. 2022. *Bloomberg - Are you a robot?* [online] Available at: <https://www.bloomberg.com/news/articles/2020-01-06/how-deepfakes-make-disinformation-more-real-than-ever-quicktake#xj4y7vzkg> [Accessed 15 June 2022].

Braunlein, F & Frerichs, L. (2019). Smart Spies: Alexa and Google Home expose users to vishing and eavesdropping. Security Research Labs. https://www.srlabs.de/bites/smart-spies#:~:text=SRLabs%20research%20found%20two%20possible,assistants%20into%20'Smart%20Spies'. Accessed 27th April 2022.

Burden, E. (2018). Husband used smart-home device to spy on wife. The Times. https://www.thetimes.co.uk/article/husband-used-smart-home-device-to-spy-on-wife-3xzcfqp3m
Braitwaite, P. (2018). Smart home tech is being turned into a tool for domestic abuse. WIRED. https://www.wired.co.uk/article/internet-of-things-smart-home-domestic-abuse Accessed 25th April 2022.

Cambridge Dictionary https://dictionary.cambridge.org/dictionary/english/trust Accessed 20th February 2022

Contact Martijn Kösters Partner mkosters@deloitte.nl, "IPA versus RPA – What's the Difference" (*Deloitte Netherlands* May 6, 2022) <https://www2.deloitte.com/nl/nl/pages/tax/articles/bps-ipa-versus-rpa-whats-the-difference.html> accessed June 16, 2022

Crawford, K. (2021) Time to Regulate AI that Interprets Human Emotions. Blog Post entry Nature's World. https://www.nature.com/articles/d41586-021-00868-5 Accessed 24th May 2022.

Defining AI | One Hundred Year Study on Artificial https://ai100.stanford.edu/2016-report/section-i-what-artificial-intelligence/defining-ai

Education, I., 2022. *What is Natural Language Processing?* [online] Ibm.com. Available at: <https://www.ibm.com/cloud/learn/natural-language-processing> [Accessed 15 June 2022].

Eric Horvitz, "Defining Ai" (*One Hundred Year Study on Artificial Intelligence (AI100)* <https://ai100.stanford.edu/2016-report/section-i-what-artificial-intelligence/defining-ai> accessed July 12, 2022

Evans M, "Woman Accessed Ex-Partner's Alexa to Torment His New Girlfriend" (*The Telegraph*August 17, 2021) <https://www.telegraph.co.uk/news/2021/08/17/woman-accessed-ex-partners-alexa-device-tormented-new-partner/> accessed July 11, 2022

GAN Integrity. 2022. *What is Code of Conduct? | Definition | GAN Integrity*. [online] Available at: <https://www.ganintegrity.com/compliance-glossary/code-of-conduct/> [Accessed 13 June 2022].

Hasselbalch AG, "A Framework for a Data Interest Analysis of Artificial Intelligence · Dataetisk Tænkehandletank" (*Dataetisk Tænkehandletank*August 17, 2021) <https://dataethics.eu/a-framework-for-a-data-interest-analysis-of-artificial-intelligence/> accessed July 13, 2022

Kim, E. (2018). Echo secretly recorded a family's Conversation and Sent it to a Random Person on their Contact List. CNBC. https://www.cnbc.com/2018/05/24/amazon-echo-recorded-conversation-sent-to-random-person-report.html Accessed 27th April 2022

Loi, M. Spielkamp, M. (2021). Towards accountability in the use of Artificial Intelligence for Public Administrations. Algorithm Watch. https://algorithmwatch.org/en/wp-content/uploads/2021/05/Accountability-in-the-use-of-AI-for-Public-Administrations-AlgorithmWatch-2021.pdf Accessed 14th January 2022

Merriam Webster Dictionary. https://www.merriam-webster.com/dictionary/trust Accessed 7th December 2021

Morrison S, "The Case against Smart Baby Tech" (*Vox*February 26, 2020) <https://www.vox.com/recode/2020/2/26/21152920/ibaby-hacking-smart-baby-monitors-bitdefender#:~:text=A%20Seattle%20couple%20reported%20last,threatened%20to%20kidnap%20the%20baby.> accessed July 11, 2022

Murph D, "IPhone 4S Hands-on!" (*Engadget*May 13, 2021) <https://www.engadget.com/2011/10/04/iphone-4s-hands-on/> accessed June 16, 2022

Mutchler A, "Voice Assistant Timeline: A Short History of the Voice Revolution" (*Voicebot.ai*March 26, 2021) <https://voicebot.ai/2017/07/14/timeline-voice-assistants-short-history-voice-revolution/> accessed June 16, 2022

Newman LH, "Turning an Amazon Echo into a Spy Device Only Took Some Clever Coding" (*Wired* April 25, 2018) https://www.wired.com/story/amazon-echo-alexa-skill-spying accessed December 12, 2021

NY Post. (2016). Toddler Asks Amazon's Alexa to Play Song but Gets Porn Instead. https://nypost.com/2016/12/30/toddler-asks-amazons-alexa-to-play-song-but-gets-porn-instead/ Accessed 27th April 2022.

Nytimes.com. 2022. *The Secretive Company That Might End Privacy as We Know it (Published 2020)*. [online] Available at: <https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html> [Accessed 15 June 2022].

Person, "Top 10 AI-Powered Virtual Assistant Companies" (*AI Magazine*April 26, 2022) <https://aimagazine.com/ai-applications/top-10-ai-powered-virtual-assistant-companies> accessed July 12, 2022

Person, "Top 10 AI-Powered Virtual Assistant Companies" (*AI Magazine*April 26, 2022) <https://aimagazine.com/ai-applications/top-10-ai-powered-virtual-assistant-companies> accessed July 12, 2022

Research and Markets, "European Conversational AI Market 2021 - 2027: German Market Dominated in 2020 and Is Forecast to Reach $202.8 Million by 2027" (*GlobeNewswire News Room*December 13, 2021) <https://www.globenewswire.com/news-release/2021/12/13/2350501/28124/en/European-Conversational-AI-Market-2021-2027-

German-Market-Dominated-in-2020-and-is-Forecast-to-Reach-202-8-million-by-2027.html> accessed July 12, 2022

Responsible Innovation Project. Can We Trust AI When We Can't Even Trust Ourselves (2021) https://responsibleproject.com/trust-in-ai-can-we-trust-ai-when-we-cant-trust-ourselves/#:~:text=When%20industry%20organizations%20and%20institutions,transparency%20into%20our%20intelligent%20systems. Accessed 14th January 2022.

Silver S, "A History of Voice Technology" (*Blog*) <https://info.keylimeinteractive.com/history-of-voice-technology> accessed June 16, 2022

Techcrunch.com. 2022. *TechCrunch is part of the Yahoo family of brands*. [online] Available at: <https://techcrunch.com/gallery/a-battle-royale-of-digital-assistants-the-big-5/> [Accessed 15 June 2022].

The Verge. 2022. *Google dissolves AI ethics board just one week after forming it*. [online] Available at: <https://www.theverge.com/2019/4/4/18296113/google-ai-ethics-board-ends-controversy-kay-coles-james-heritage-foundation> [Accessed 13 June 2022].

Webopedia. 2022. *What Is Conversational AI? | Webopedia*. [online] Available at: <https://www.webopedia.com/definitions/conversational-ai/> [Accessed 15 June 2022].

Webster Dictionary https://www.merriam-webster.com/dictionary/trust Accessed 20th February 2022.