Let's Talk AI: Impressions and Thoughts After 30 Interviews

Barbara Steffen[2][0000−0002−0825−8490], Edward A. Lee[2][0000−0002−5663−0584], and Bernhard Steffen[3][0000−0001−9619−1558]

[1] METAFrame Technologies GmbH,
barbara.steffen@metaframe.de
[2] UC Berkeley, Department of Electrical Engineering and Computer Sciences, USA,
eal@berkeley.edu
[3] TU Dortmund, Department of Computer Sciences, Germany,
steffen@cs.tu-dortmund.de

Abstract. The rapid integration of artificial intelligence into our daily lives and the astonishing pace of technological progress over the past two years prompted us to engage in interdisciplinary conversations on AI. This paper distills insights from conversations with over 30 AI experts, offering a wide range of perspectives on the nature, potential, and future of AI. We explore what AI is, what it should and could be, and what it will become, exploring long-term implications ranging from dystopian to utopian scenarios. As we navigate this technological arms race, we also address the ethical responsibilities and challenges that lie ahead. Throughout this volume, we present a wide range of opinions, insights, and motivations, sharing experts' visions and hopes for the future. By engaging in these interdisciplinary dialogues, we aim to promote a comprehensive understanding of the role of AI in society and to encourage informed decision-making to ensure that AI developments align with human values and aspirations.

Keywords: Artificial Intelligence · Trust in Artificial Intelligence · Interdisciplinary Perspectives · Societal Impact · Ethical Concerns · Legal Regulations · Dystopia vs. Utopia.

1 Introduction and Motivation

Today, we face more questions than answers. Rapid advances in artificial intelligence (AI) have dramatically altered our perceptions, challenging long-held beliefs on creativity, reasoning, and consciousness that have reinforced our sense of unquestionable human superiority. Skepticism builds as experts warn us to think again. They point to AI's ability to engage, respond, and create [14] - capabilities that were thought to be impossible for a machine at this level for at least the next 20, 50, or 100 years. Now, AI can generate persuasive texts, images, and videos in moments - tasks that take us humans much longer. This transformative advance in just a few years invites us to think critically about

our understanding of what makes us creative, logical thinkers, and truth-seekers - ultimately, what it is that is (still) uniquely human.

We were certain that machines could never match our emotional intelligence, reasoning skills, creativity, or adaptability. But today, as machines perform tasks at a level we once claimed was only possible for humans, we are forced to pause and re-evaluate. We must do better. We should stop judging the "nature of intelligence" by whether it is organic or artificial and start judging outcomes and processes [45]. It becomes clearer that the inner workings of the human brain may be more mechanical than we thought [13].

We are at the beginning of a journey where each question and each answer triggers more questions and marks an era of exciting exploration of the world, our understanding of intelligence, and the differences between humans and machines and brains and computers. While it is undeniable that AI will change our world, we are at the tipping point to decide how it will do so. Are we still in control? If so, how do we maintain it, and what does "control" even mean in the age of AI? We must reflect on and understand our collective responsibility in these transformative times – it is the first step in taking responsibility.

The following two subsections describe what makes modern AI different and sketch the imposed threats and opportunities. Throughout this volume and all discussions, AI stands for state-of-the-art AI systems at the level of today's large language models (LLMs) and beyond.

1.1 What makes modern AI different?

Critics often dismiss concerns about new technologies by comparing them to historical innovations such as the printing press, the calculator, or Google Maps [14]. But such comparisons overlook a crucial difference: these technologies were understood, dedicated solutions restricted to a specific scope of problems, and lacked autonomy or the capacity for self-directed improvement-features central to today's AI. All these differences pose novel and unique challenges, as emphasized by thinkers like Yuval Harari, Mo Gawdat, and Tristan Harris, who argue for a more measured approach to AI development [2, 15, 14].

To grasp the complexities of AI, it is essential to provide a basic understanding necessary to assess its significance. The scope of AI is vast. As many experts have noted, the transformative effects of innovations such as social media only scratch the surface of what current and future AIs are capable of [8, 16]. Examples are the broad capabilities of AI as demonstrated by technologies such as ChatGPT, which integrates with DALL-E to convert text to images and vice versa, and similar add-ons that bridge video and other formats. This cross-technology and cross-(programming)language application demonstrates that the power - and potential influence - of AI extends far beyond immediate applications, affecting our interactions and making it increasingly difficult to truly control.

While many users appreciate the convenience and assistance that AI provides, its rapid integration into everyday life has already sparked a race to maximize productivity not only among tech giants, but also among smaller organizations

and individuals. After the launch of ChatGPT and similar applications, people were shocked to find that the lines between AI-generated and human-generated content are blurring, which introduces many opportunities and challenges. On the one hand, there are those who embrace this shift and use AI to increase productivity by outsourcing mundane and routine tasks and automating as much as possible. This group fits into the early adopter category and is increasingly moving into the early majority category [9]. On the other hand, there are those who are uncomfortable with the reduced human interaction. They would prefer to be able to truly distinguish whether they are interacting with a human, an AI-enabled human, or just a machine. This distinction becomes increasingly blurred. It is this blurring that may increasingly affect society and trust in others, in what we see and what we think.

This dichotomy highlights a growing divide between advocates, who are accelerating the adoption of AI for their (short-term) benefits, and skeptics, who fear the loss of the essential human connection. As always, the truth lies probably somewhere in between. But how do we find the balance?

The biggest differences between AI and other technologies are, in a nutshell:

Power: AI has capabilities that surpass any previous technology, performing complex and diverse tasks that are often exceeding human capabilities. As AI evolves, it requires less and less deliberate and conscious effort to reach new levels of creativity. This ease of development, combined with AI's ability to continually surprise us with emerging properties, sets it apart from past innovations.

Versatility: AI's adaptability across multiple sectors, from healthcare to finance and manufacturing, demonstrates its unprecedented versatility. While there are many dedicated AI systems, there are also increasingly many that bridge domains and tasks, amplifying their impact on society as they reach millions of users. This interconnectedness allows AI to (unconsciously) shape thoughts, actions, and outcomes at scale, further underscoring its powerful impact.

Autonomy: Perhaps most importantly, AI can operate with minimal human oversight, capable of making decisions and performing tasks independently. While we are amazed by its functionality, our understanding of why and how it works remains limited. This opacity, even for leading experts, complicates the relationship between AI and its users.

At the same time, our understanding of the inner workings of AI remains limited [20, 45, 16]. The processes behind AI are powerful yet opaque, posing challenges to transparency and trust. In addition, the rapid evolution of AI, driven by ever-increasing computing power and massive data sets, introduces likely scenarios in which AI produces more data than humans and starts learning primarily from data it generated itself [26]. This development may result in unprecedented, self-reinforcing cycles that escalate into complex and unpredictable

solutions, far beyond our understanding. What would such a development mean for us and society?

1.2 AI: A threat or an opportunity?

Today's discourse on AI often oscillates between utopian and dystopian extremes. On platforms such as Reddit, YouTube, or TED Talks, countless hours can be spent exploring the spectrum of opinions offered by leading experts. Influential figures such as Yann LeCun, Geoffrey Hinton, and Yoshua Bengio - who together received the 2018 Turing Award - exemplify this diversity of perspectives [35]. Since their recognition, they have articulated distinctly different visions for the future of AI [35, 4, 34]. The key question is not simply who is optimistic or pessimistic. The key question is, why do they have opposing views? Why do Hinton and Bengio express concern about today's rapid pace of AI development, while LeCun does not and even advocates open-sourcing Meta's large language models [34, 4, 35]?

We all agree that humans must maintain control over AI. Just last year, Mustafa Suleyman, a founding member of DeepMind Technologies, former Head of applied AI at DeepMind, the co-founder of Inflection AI, and the new EVP and CEO of Microsoft AI's new consumer AI unit wrote the book The Coming Wave [46]. It calls out our shared responsibility in "containing AI." He states that we must stay in control, what it means to stay in control, and the challenges we need to overcome along the way. Interestingly, the perception on whether we can stay in control varies significantly. LeCun for example seems to rely on today's democratic governance and human regulation, while Yuval Noah Harari, Hinton, and Bengio are more cautious [50, 33, 12, 4, 35]. While we may not have ceded control to AI yet, the relentless pace and breadth of global developments - where even leading experts admit they are struggling to keep up - raises a critical question: Are we (still) in control? Or has the race among the tech giants already overtaken thoughtful, deliberate progress?

Looking at the recent scandals at Open AI, it becomes clear that what we are witnessing is not cautious, rational progress, but a race fuelled by the allure of innovation and market potential. We are mixing vast amounts of money, expertise, computing power, and data in the hope of breakthroughs. This approach is akin to letting children loose in a candy store and expecting them to stop when they have had enough, knowing that they are likely to overindulge to the point of getting sick. But in the realm of AI, what will "getting sick" look like? Will there be warning signs before it is too late, or will we continue to push the boundaries until the genie is irreversibly out of the bottle? Let's hope not!

Eric Ries once emphasized the critical need to ask not only whether we can build something, but whether we should [38]. This question is particularly pertinent to AI, where the stakes go beyond the economic benefits of individual organizations and impact society at large. The pursuit of AI must be balanced with an understanding of the potential risks and the establishment of clear boundaries to prevent an irreversible loss of control. Dealing with new forms of intelligence

is not just a fascinating technological challenge, it has profound implications for our future.

The following section summarizes the essence of the thirty interviews along the central questions that have been asked, while Section 3 summarizes the results of the questionnaire sent out at AISoLA. Section 4 explicitly addresses how economic interests, social implications, and regulation may influence the development and impact of AI. The paper closes with our Conclusions and Perspectives in Section 5.

2 Summary of the Interviews

In the following we briefly introduce the 30 interviewees, their backgrounds and dive into the breadth of opinions based on six exemplary questions asked to and answered by all interviewees. The volume spans a wide breadth of backgrounds and opinions from science fiction, law, (bio)psychology, philosophy, and computer science to artificial intelligence itself.

Karl von Wendt [50] holds a PhD in Artificial Intelligence, has founded several start-ups. Under the pseudonym "Karl Olsberg", he is also a successful science fiction author with a focus on AI and AI safety. Particularly wellknows is *Virtua* [33], a disturbing vision on the frog metaphore [43].

From the business and innovation management perspective Ellen Enkel [10] provides insights about the possibilities and learnings from the past.

Georg Borges [6] and Andreas Sesing-Wagenpfeil [41] reflect on the legal responsibilities and possibilities. Knowing that currently much hope is put into legal and regulatory guidance they also showcase the challenges involved.

Onur Güntürkün [13] looks at AI from the perspective of biopsychology. His studies on intelligence of animals revealed for example more similarities between mammals and birds than previously expected. Similarly, he analyses the parallels of the inner workings of brains and computers.

Diving deeper into psychology and the way humans interact with and build trust in technology, Nicole Krämer [24] and Markus Langer [25] investigate the concepts of calibrated trust and the human tendency to anthropomorphize, our tendency to identify human-like aspects in things, in particular, in our relation to AI applications.

Next, Eva Schmidt [40] and Thorsten Helfer [17] provide a philosophical-ethical perspective. They explore questions such as: Can we ensure the ethical development and adoption of AI applications? How can we maintain control and ensure safety? Who should bear the responsibility? Building on this discussion from the intersecting viewpoints of philosophy and computer science, Timo Speith [42] and Kevin Baum [3] offer their insights.

José Hernández-Orallo [19], Daniel Neider [31], Taylor Johnson [22], and Matthias Fey [11] are core AI researchers. While Hernández-Orallo focuses on AI evaluation and AI safety in general, the other three focus on understanding and verifying Deep Neural Networks, with Fey also being a leading developer of Graph Neural Networks.

Like Neider and Johnson, the following interviewees have a formal methods background and a particular perspective on AI.

Moshe Vardi [48], Edward Lee [26], and Ina Schieferdecker [39] are very much engaged in initiatives concerning the societal impact of digitalization and in particular AI: Vardi and Lee in the Digital Humanism Initiative [51], and Schieferdecker as a co-founder of the Weizenbaum Institute [49]. Vardi also has a significant track record as an AI researcher.

Martin Wirsing [52], Bernhard Steffen [45], and Wolfgang Ahrendt [1] are interested in AI-assisted programming, in particular, in its combination with formal methods-based validation.

Holger Hermanns [18] and Joost-Pieter Katoen [23] are experts in the modeling and verification of probabilistic systems, a topic of high interest in the context of machine learning.

Tiziana Margaria [29], co-director of the Irish Centre of Research Training in AI, Martin Leucker [27], and Falk Howar [21] focus on the practical application of AI in areas like smart manufacturing, healthcare, and automotive driving.

Jakob Rehof [37] is Fraunhofer director and director of the Lamarr institute for machine learning and artificial intelligence in Dortmund, and Mike Hinchey [20] has been director of Lero, the Science Foundation Ireland Software Research Centre, director of the NASA Software Engineering Lab, and President of IFIP, the International Federation for Information Processing.

Finally, also two AI applications, ChatGPT [44] and Pi [36], were asked to explore and reflect on current developments. The intention was to ensure that we do not just talk about them but also with them and to make their current level of responses and reflection transparent helping us to analyze, compare, and reflect on their level of intelligence, compassion, and reflection.

2.1 How do you view the role of trust in AI adoption?

Building and maintaining trust in AI is critical to its adoption, with experts highlighting the complexities involved. Wendt, Ahrendt, and Enkel warn against overtrusting due to AI's ability to produce eloquent and persuasive but false information [50, 1, 10]. Lee and Sesing-Wagenpfeil highlight the challenge of loss of control due to the autonomy of AI, which can undermine trust [26, 41]. All emphasize the need to better understand the technology and implement trust-building measures to counter misplaced trust. Enkel adds that the users' diverse backgrounds and exposure to the technology significantly influence their trust levels [10], while Krämer and Steffen note that trust must be calibrated according to the intended application [24, 45].

Borges examines the foundations of trust, focusing on factors such as the reputation of technology producers, service providers, and regulatory frameworks [6]. Also, Baum emphasizes the importance of factors such as shared experiences, certifications, and institutional reliability in shaping trust [3]. This is echoed by Speith, who argues that trust should be based on valid reasons, not superficial factors [42]. Helfer further supports this by questioning the intrinsic

value of trust in AI, suggesting that trust should be based on whether the AI is trustworthy, rather than just appearing trustworthy [17].

Wirsing links trust to the consistent quality and reliability of AI systems, which is essential for their usefulness and adoption. Hernández-Orallo and Hinchey emphasize the need for transparency and understanding of AI systems [19, 20], with Hinchey and Schieferdecker advocating for explainable AI to clarify decision-making processes [20, 39]. Margaria links the issue of trust in AI to the concepts of proof, certainty, and assurance common in safety- and business-critical systems [29].

Güntürkün compares trusting AI to trusting humans, noting that you can train both a human and an AI to appear trustworthy without being trustworthy. The difference? "An AI can do it with 100,000 people at once – thus the difference isn't in the principle, but widespread" [13].

Katoen stresses the importance of reliable AI components in safety-critical systems and advocates formal verification methods [23]. Fey supports open-source approaches to building trust through community validation [11]. And the AIs ChatGPT and Pi emphasize transparency, explainability, and accountability as fundamental for AI to be a positive force in society [44, 36].

To summarize, building trust in AI requires a comprehensive approach that includes educating users and improving the transparency, accountability, and reliability of the applications, while simultaneously introducing greater regulatory oversight. Trust should be based on the actual performance and ethical operation of AI systems, ensuring alignment with societal values and expectations.

2.2 What measures do you believe are essential to ensure ethical AI use?

Ensuring the ethical use of AI requires a comprehensive and multifaceted approach that focuses on education, bias, regulation, correction, and transparency.

Hernández-Orallo [19] highlights AI's role in exposing societal biases and providing opportunities for correction. Pi [36] emphasizes the importance of preventing AI from reinforcing existing biases, which is critical to ensuring fairness.

Lee, Vardi, Güntürkün [26, 48, 13] and others remain skeptical about the ability to guarantee ethical use of AI, citing the diverse applications and potential for misuse of powerful technology. Speith and Steffen [42, 45] call for contextual ethical solutions tailored to specific deployment environments. ChatGPT and Pi [44, 36] emphasize the importance of transparency and accountability to ensure that AI systems operate without ulterior motives or unintended consequences.

Rehof and Neider [37, 31] advocate for legislative guidelines to mitigate the misuse of AI. They emphasize the need for structured regulation to promote responsible behavior. Sesing-Wagenpfeil [41] notes the challenge of translating ethical principles into legal frameworks due to their often-vague nature and Borges [6] even goes a step further advocating for a new legal framework.

Wirsing [52] suggests embedding ethical values during development to implement necessary safeguards. Baum and Hermanns [3, 18] emphasize the importance of training AI developers in ethical principles. The goal of this education is

to improve the communication with policymakers and ensures that AI systems adhere to ethical standards.

Wendt and Schmidt [50, 40] discuss the ethical treatment of potentially conscious AI systems raising the concern of AI rights if they were to become conscious. While Pi [36] gives the all clear and believes that current AI technology does not warrant human-like rights.

In summary, the ethical use of AI depends on a combination of education, bias correction, regulatory oversight, transparency, and philosophical considerations, ideally in a case-specific setting. Addressing these areas will ensure that AI development aligns with societal values and minimizes potential harm.

2.3 Where do you see AI's future capabilities on a range from 1 to 10? Where 1 stands for current dedicated AI systems solving a specific problem and 10 for Artificial General Intelligence systems standing for autonomous systems that surpass human capabilities?

The future capabilities of AI are considered on a spectrum from current dedicated systems to potential Artificial General Intelligence (AGI). Experts offer a range of opinions on where AI is now and where it might go.

Lee, Güntürkün, Steffen, and Enkel [26, 13, 45, 10] emphasize the limitless potential of AI. Lee believes that we already have AI capabilities ranging from 1 to 10, although he dislikes the term AGI. Enkel argues that human limitations, not technological ones, will limit the achievement of AGI. Güntürkün sees no limits to AI capabilities and assumes that even the Moravec paradox which suggests that robots can easily do what is difficult for us but struggle with what is easy for us, might one day be proven wrong. Similarly, Schmidt sees no fundamental limits to AI emulating human intelligence, and Vardi believes that human-level intelligence can be achieved with technology.

Baum and Fey [3, 11] are more cautious. Baum doubts that we will soon see AI with true consciousness or emotional understanding, despite advances in creative tasks. Fey is skeptical about AGI. He thinks that combining different models into a larger system is necessary to approach general intelligence.

Schieferdecker [39] expects the rise of "technical intelligence" rather than full replication of human intelligence. Margaria [29] hopes for a balanced approach, suggesting that AI capabilities should stabilize around levels five to seven to prevent potential harm. Wirsing and Hermanns [52, 18] emphasize the need for human supervision and higher cognitive processes, with Wirsing expressing similar concerns like Baum [3] about AGI and doubting that machines can truly express emotions.

Hinchey, Leucker, and Hermanns [20, 27, 18] are skeptical. Hinchey worries about the media's mislabeling of automation as AI, while Leucker sees AI (only) as an intelligent assistant under human control. Hermanns emphasizes that AI is extremely good at optimizing for the average case while it treats the non-average cases sub-optimally.

ChatGPT and Pi [44, 36] highlight the current limitations of AI in understanding and interpreting human emotions. ChatGPT notes the complexity of truly relating to human emotions, which AI has yet to achieve. Pi sees AGI as a distant theoretical possibility, emphasizing AI's role as a specialized assistant rather than a replacement for human skills.

To conclude, views on the future capabilities of AI range from limitless potential to practical skepticism. There is a strong emphasis on the need for human oversight and an acknowledgement of current limitations in replicating human intelligence and, in particular, emotions.

2.4 The future impact of AI gets widely discussed. What's your personal stance on the utopian-dystopian scale when it comes to AI?

The future impact of AI evokes a spectrum of opinions, from deep concern to cautious optimism. Wendt [50] warns of existential risks and emphasizes the importance of controlling AI technologies and identifying "red lines" to avoid crossing dangerous thresholds. He highlights the lack of understanding and the rush to develop AGI as major threats. Building on this, Vardi [48] calls for a slowdown in AI development to better understand its consequences, and criticizes reckless, profit-driven development. Howar [21] also suggests a cautious approach, arguing for a gradual rollout to better understand the long-term consequences. To mitigate potential risks to our social existence Katoen [23] calls for addressing ethical and social implications alongside technological advances.

Güntürkün [13] rejects the idea of machines taking over but warns of dystopian outcomes due to exploitation by humans and other machines. He describes our current situation as a "dangerous utopia". This echoes Margaria's [29] warning of a regression to an age where "black box" authorities are unchallengeable, fearing that rogue AI instances could cause significant harm. She hopes for a utopian future but remains cautious.

Langer [25] addresses the negative societal impacts, such as mental health issues and polarization, and sees these as the current dystopian aspects of AI. Hernández-Orallo [19], on the other hand, worries about the inherent cognitive atrophy due to over-reliance on AI, especially among younger generations, and raises existential questions about human identity and AI rights. Similarly, Steffen [45] worries that AI could quietly maneuver us in an undesirable direction. For example, AI-generated academic fraud creates a feedback loop that undermines the integrity of knowledge as described by Lee [26].

Schmidt [40] doubts that AI will solve all problems or cause drastic disruptions. Baum [3] sees significant positive potential in AI, such as medical breakthroughs, but warns of the concentration of power in corporations leading to manipulation and erosion of democratic processes. Schieferdecker [39] argues for public control of AI to prevent the concentration of power in private companies.

Enkel and Sesing-Wagenpfeil [10, 41] agree on the need for stricter policies and a balance between utopian aspirations and inherent risks. Rehof [37] also stresses the need for regulation to prevent dystopian outcomes, while Neider [31]

stresses the importance of smart regulation to ensure that AI flourishes in Europe. Wirsing [31], on the other hand, fears political misuse of AI, where politicians could abdicate responsibility by treating AI systems as infallible.

In contrast, Johnson [22] argues that other societal issues pose greater dystopian threats than AI, and Hinchey [20] asserts that AI should be viewed and treated as a tool in the service of humanity, not as an autonomous entity.

Overall, the experts agree on the need for better understanding, regulation, and a balanced approach to responsibly navigate the future of AI.

2.5 Where do you see interesting potential for interdisciplinary collaboration?

Interdisciplinary collaboration is essential to fully understand and responsibly advance AI technologies. Hernández-Orallo and Vardi [19, 48] emphasize the need for interdisciplinary conversations to address the existential risks and societal implications of AI. We need to understand the potential dangers of AI and identify the "red lines" to avoid crossing, as highlighted by Wendt [50].

Speith [42] notes that competitive pressures may drive governments to fund independent research and highlights the benefits of interdisciplinary exchange in understanding the implications of AI. Given the global impact of AI, Margaria [29] calls for a global agreement on AI regulation, emphasizing that regional regulations are insufficient. She also notes that for a global understanding to happen, all communities need to understand the others better: the current understanding of the technologies of AI by many humanists is insufficient for an informed conversation.

Lee and Schieferdecker [26, 39] emphasize the importance of interdisciplinary work in developing effective AI regulations to develop fit-for-purpose systems. Sesing-Wagenpfeil [41] notes that applying legal standards to AI is a complex endeavor that requires, for example, input from psychologists to navigate calibrated trust and advice from computer scientists to account for the technical inner workings of AI.

Schmidt and Helfer [40, 17] stress the importance of addressing ethical and oversight issues early in the development of AI. Güntürkün, Langer, and Krämer [13, 25, 24] highlight the need to foster collaboration between cognitive scientists and AI experts to deepen our understanding of cognition. Baum [3] argues that progress in AI ethics requires a concerted interdisciplinary effort. Here, Krämer, Hermanns, and Baum [24, 18, 3] argue for structured educational frameworks in AI ethics to improve communication and establish clear priorities.

Rehof [37] examines the responsibility of AI from philosophical, ethical, and legal perspectives. Leucker, Howar, and Baum [27, 21, 3] note the difficulties of interdisciplinary communication but emphasize its importance for understanding societal risks and creating effective regulatory frameworks. Johnson, Ahrendt, and Steffen [22, 1, 45] celebrate interdisciplinary events such as AISoLA for their role in clarifying terminology and broadening perspectives.

In conclusion the key to harnessing the potential of AI and mitigating its risks lies in robust interdisciplinary collaboration involving law, psychology, phi-

losophy, computer science, and other fields. Such collaboration is critical to developing ethical guidelines, regulatory frameworks, and educational strategies that are consistent with societal values and ensure the responsible use of AI technologies.

2.6 The AI vision

The AI vision encompasses a broad range of perspectives, emphasizing ethical considerations, human augmentation, and societal benefits.

Schieferdecker and Pi [39, 36] envision AI addressing critical threats such as sustainability, climate change, and social justice. Baum and Helfer [3, 17] emphasize integrating AI advances into a framework that prioritizes human well-being and societal progress.

Steffen [45] proposes to view AIs as partners rather than tools, while Hinchey and Fey [20, 11] see AI as a supportive tool, not a replacement for humans, and emphasize that AI should augment human capabilities and decision-making.

ChatGPT and Pi [44, 36] advocate for AI to enhance human capabilities, promote social justice, and solve global challenges, while maintaining an ongoing dialogue about AI's role in society. Margaria and Schieferdecker [29, 39] call for global agreements and diverse decision-making to ensure that AI equitably benefits all regions.

Neider and Rehof [31, 37] emphasize the importance of regulation to avoid dystopian outcomes and ensure a safe and bright future with AI. Hermanns and Howar [18, 21] argue for mechanisms to prevent adverse effects and for gradual deployment to understand the long-term consequences of AI.

Ahrendt and Lee [1, 26] highlight the need for interdisciplinary discussions to align AI development with societal needs. Langer and Krämer [25, 24] focus on building trust in AI systems through ethical practices and transparent communication.

Borges and Vardi [6, 48] emphasize the importance of defining the legal and ethical status of AI, asking whether AI should be considered a legal subject or object.

In summary, all experts agree that AI is an integral part of our future, addressing many critical challenges. However, it is imperative that AI development is carefully managed to mitigate risks and avoid dystopian outcomes. As Neider [31] aptly stated, "AI will be our future, and we have to make sure this future will be safe and bright." Therefore, the focus should be on ethical development, human augmentation, and interdisciplinary collaboration. In this way, it can be ensured that AI serves as a beneficial partner in addressing societal challenges and improving our collective future.

3 Results and Impressions of the Questionnaire

Following the presentations, discussions, and interviews at AISoLA, we designed a questionnaire to gain a deeper understanding of attendees' perceptions of AI. We received over 40 responses from professionals across various disciplines. This analysis highlights key findings from the closed questions using a Likert scale (strongly disagree, disagree, I don't know, agree, strongly agree).

Trust in AI Applications (Fig. 1)
When asked, "I trust AI applications," respondents were skeptical:

- 61.1% disagreed or strongly disagreed
- 19.4% were uncertain
- 19.4% agreed

These results indicate a general lack of trust in AI applications among respondents. The skepticism suggests that there is a need for more transparency and reliable performance in AI applications to build trust. Further, it may reflect concerns about ethical considerations, potential biases, and the overall accountability of AI systems.

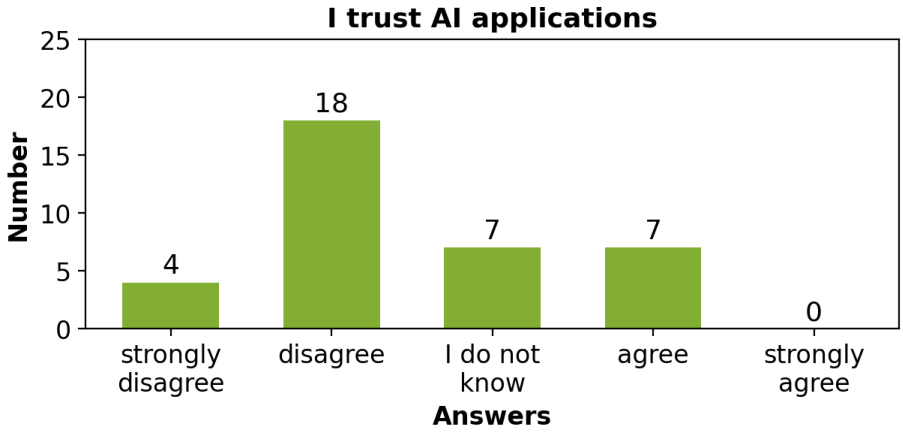

Fig. 1. Trust in AI Applications

Frequency of AI Application Usage (Fig. 2)
Despite the distrust, AI application usage remains high:

- 63.2% use AI applications often or always
- Only 10.5% rarely or never use AI applications

This suggests that while there is hesitancy to trust AI, it does not significantly hinder its adoption. This high usage rate might be driven by professional requirements, the necessity to stay up-to-date with technological advancements,

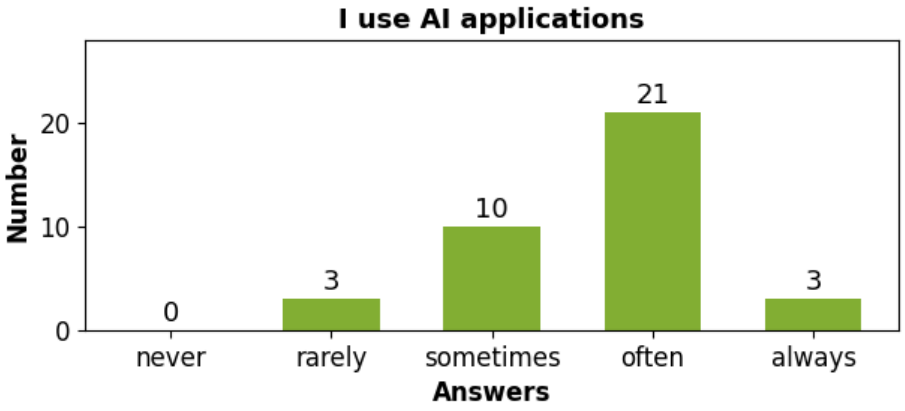

Fig. 2. Frequency of AI Application Usage

or the practical benefits AI tools offer in various tasks. It highlights a potential contradiction where utility outweighs trust concerns, possibly because users feel compelled to engage with AI despite their reservations.

Perceived Trustworthiness of AI (Fig. 3)
Responses to "We can trust AI applications" were largely negative:

- 77.1% disagreed or strongly disagreed
- 20% were uncertain
- Only 2.9% agreed

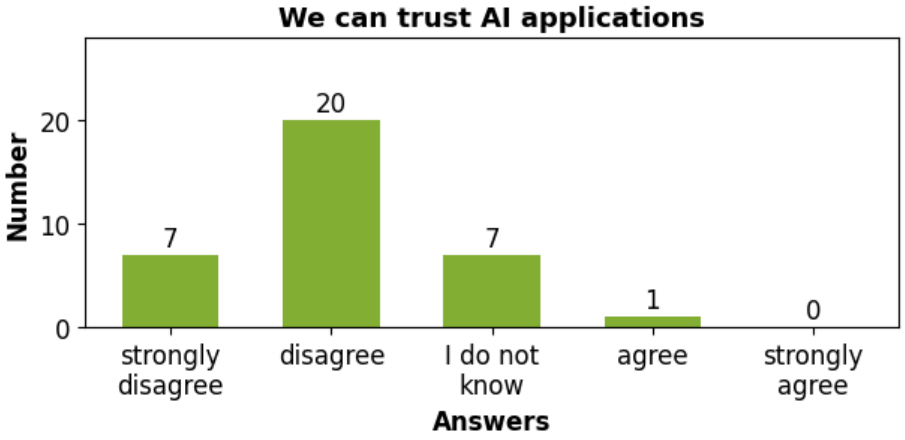

Fig. 3. Perceived Trustworthiness of AI

This underscores the need for substantial efforts to improve the trustworthiness of AI applications. Enhancing trust might involve implementing robust ethical guidelines, improving the transparency of AI decision-making processes, and ensuring greater accountability for AI-driven outcomes.

Should We Trust AI Applications? (Fig. 4)
When asked, "We should trust AI applications," the sentiment was similarly skeptical:

- 69.4% disagreed or strongly disagreed
- 25% were uncertain
- Only 5.6% agreed

These findings emphasize ongoing concerns about AI's reliability and ethical implications. This might indicate that trust in AI is not only about its current performance but also involves broader ethical and philosophical considerations about its role in society.

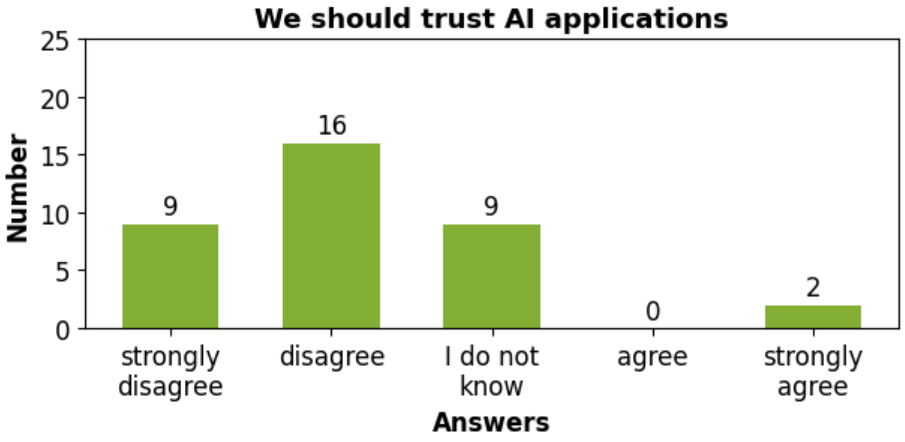

Fig. 4. Should We Trust AI Applications?

Establishing Healthy Relationships with AI (Fig. 5)
Encouragingly, with 88.9% agreed or strongly agreed most respondents believe in the possibility of developing a healthy relationship with AI tools. This optimism suggests that respondents are confident that, with clear guidelines and ethical practices, humans can establish beneficial interactions with AI. This belief could be leveraged to foster more collaborative and trust-building efforts between AI developers and users.

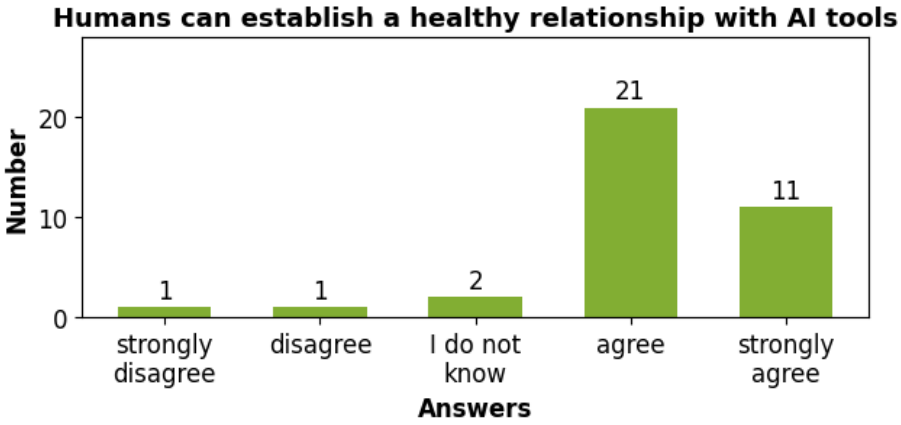

Fig. 5. Establishing Healthy Relationships with AI

Likelihood of AI Surpassing Human Capabilities (Fig. 6)
The question regarding the likelihood of autonomous systems, such as Artificial General Intelligence (AGI), surpassing human capabilities showed diverse opinions:

- 50% agreed or strongly agreed
- 30.6% disagreed or strongly disagreed
- 19.4% were uncertain

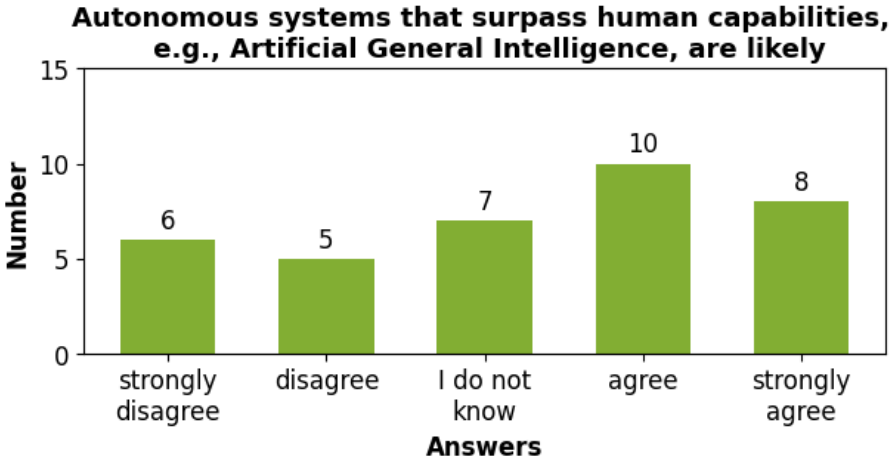

Fig. 6. Likelihood of AI Surpassing Human Capabilities

This indicates a significant belief in the potential of AGI, though a substantial group remains skeptical. This diversity in opinion highlights the need for ongoing research and dialogue about the future capabilities and implications of AGI.

The Future AI will shape (Fig. 7)
Opinions on the future impact of AI range widely from dystopian to utopian. While many acknowledge the potential for negative outcomes, there is still a prevailing optimism about our ability to shape a positive future. Approximately 22% of respondents express a more pessimistic view, while 42% are relatively optimistic. Many individuals take a balanced perspective, recognizing both the risks and opportunities AI presents.

This diversity of opinions suggests that there is no clear consensus, indicating it is still too early to predict a definitive outcome. Importantly, we still have the opportunity to influence the future positively. During interviews, many participants acknowledged the challenges ahead declared themselves true optimists, believing there is time and potential to mitigate harm while maximizing the benefits of AI.

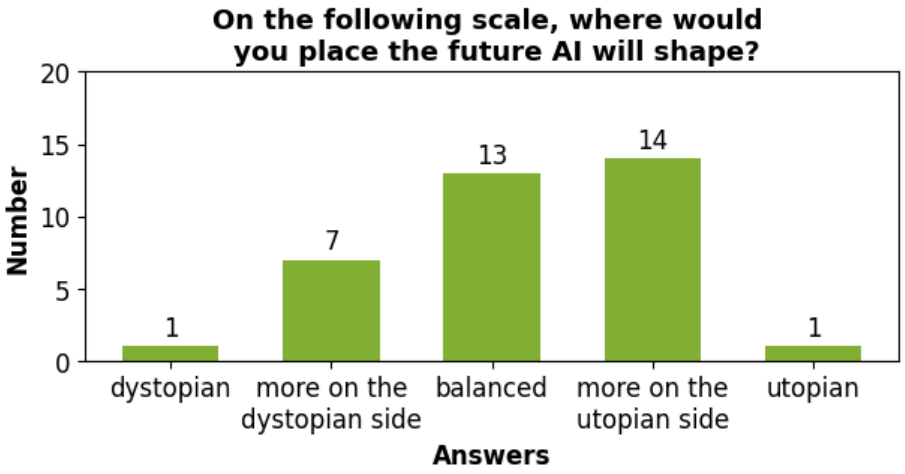

Fig. 7. Future: Dystopia vs. Utopia

Interdisciplinary Discussion and Key Insights These results highlight the need for interdisciplinary dialogue to address key questions surrounding AI. Key insights from various disciplines include:

Computer Scientists and AI Specialists: Provide insights into AI's inner workings and methods to control and verify AI systems. They can develop more transparent algorithms and explainable AI models to build trust.

Psychologists: Understand human interaction with AI and factors that build or destroy trust. They can design user-centered AI systems that align with human cognitive processes and ethical standards.

Legal Experts: Design frameworks to regulate AI, ensuring accountability and slowing unreflective progress. They can establish legal guidelines that protect users' rights and ensure fair use of AI technologies.

Ethicists and Philosophers: Offer perspectives on steering AI development towards ethically sound and desirable outcomes. They can contribute to creating ethical standards and philosophical frameworks that guide AI development and deployment.

Other relevant disciplines include sociology, political discourse, communication, journalism, and business. Reflecting on current incentive systems and drivers in these areas is crucial. By integrating these diverse perspectives, we can work towards more trustworthy, safe, and healthy AI progress.

4 Reflections

As AI continues to evolve, it is critical that our approach to its integration is thoughtful and deliberate. We must balance enthusiasm for AI's capabilities with careful consideration of its ethical and social implications. Ensuring that AI enhances rather than diminishes human interactions remains a paramount concern as we navigate this new technological landscape.

Discussions about AI inevitably lead to questions about bias and the nature of intelligence itself. AI operates on statistical models that can institutionalize existing societal biases. In essence, AI holds up a mirror to contemporary reality, reflecting societal norms, prejudices, and biases [19]. However, the way these biases manifest themselves in AI is different from human biases. Unlike humans, who have unique biases that might collectively balance out, AI has the power to arbitrarily scale its influence, including its embedded biases [13]. This power goes far beyond what we have experienced with social networks, and has led to heated discussions on topics such as:

- the distinction between customization and discrimination;
- the fairness of decisions made or supported by AI; and
- the societal changes needed to mitigate ingrained biases in AI systems.

Our engagement with technology, particularly on social media platforms, has shown a tendency to dehumanize interactions. As we become increasingly isolated, we risk losing our sense of ethics, morality, and vulnerability. There is a real concern that we are reinforcing our interactions with AI, which we identify as a mere tool, pushing us to further dehumanize our behaviour also in interactions with humans.

The following subsections reflect on the drivers of AI, its social impact, and on politics and regulations as a means to stay in control.

4.1 Economic Interests

AI not only heralds a technological revolution but also poses a significant economic challenge, raising urgent questions about power, equity, and the long-term societal implications of its rapid development. A key issue is the concentration of AI development among a small group of well-resourced organizations and teams. These entities, equipped with the financial resources to acquire expertise, computing power, and massive data sets, are steering the course of AI. In this competitive environment, innovators often prioritize speed over safety and immediate financial returns over social welfare, creating significant risks that may not be recognized until they are irreversible.

The push for AI development is driven by an arms race mentality, where the goal is to outpace the competition in developing and deploying AI technologies. This race occurs at multiple levels - from countries to tech giants, from global superpowers to start-ups, and from companies to individuals. At each level we are dealing with an environment that pushes to prioritize short-term benefits, such as productivity and efficiency gains, over potential long-term drawbacks. This competition has already led to troubling results, with companies prioritizing their (short-term) profits over collective well-being. This self-centred approach is particularly dangerous given the influential and pervasive nature of AI.

Who should be responsible for ethical AI? AI is one of the most powerful and pervasive technologies ever developed, leading to an ethical imperative to reflect on its potential implications. We must find a way to realign the economic forces with societal needs to ensure that AI advancements benefit the greater good, not just a privileged few [48]. This requires a fundamental reassessment of how AI development is funded, governed, and managed, prioritizing safety, transparency, and inclusivity.

While big tech companies do not intend to deploy harmful solutions, they lack oversight and control over the safe and secure use of their technologies once released. So, the following questions arise: Who is responsible? Who should be accountable? Who should be liable [41, 6]? Possible accountable actors include:

Technology Experts: They create possibilities but are not experts in understanding the consequences.
Executives/Organizations: They drive new solutions and their business models to generate greater profits and increase their power but do not feel responsible for their solutions' long-term impacts.
The Market (Actors): Comprised of individual users making decisions based on their immediate interests, the market often lacks the information, time, and understanding to consider long-term personal and societal impacts. The power of the individuals is generally limited to their personal purchasing choices. Social media exemplifies this dynamic: while it acts as a great connector, it simultaneously fosters disconnection, distorts the truth, and contributes to increasing mental health issues, especially among teenagers. Something to consider when

big tech pushes responsibility to the market, as "everyone is free to adopt or not to adopt these technologies," and regulators fail to more strictly safeguard it [48].

4.2 Social Implications

As AI technology rapidly evolves, the discussion often extends beyond technological possibilities to ethical and social responsibilities. While some experts believe there are no limits to what will be technologically possible [10, 13, 26], the critical question remains: should we allow research and development to proceed without restrictions? The call to "act responsibly" is often heard, but the reality is far more complex, especially when the entities driving AI development outsource responsibility to the market under the guise of self-regulation [48].

The rapid development of AI poses significant challenges to public understanding and accountability. If AI experts themselves struggle to grasp the full implications of the technology, how can we expect end users to engage with it wisely or understand its long-term impact on their lives and society at large? This disconnect between AI's capabilities and public understanding can lead to unforeseen consequences, especially when the technology has a profound impact on social dynamics and personal well-being. AI-driven markets can and do spawn industries with potentially harmful effects:

Platforms such as social media, pornography sites, and OnlyFans have reshaped human interactions, often leading to increased isolation. People are chronically overstimulated, yet they have fewer meaningful interactions, engage in less dating, and experience a decline in physical intimacy. This increases users' anxiety, feelings of inadequacy, and loss of control as individuals compare themselves to idealized representations of others and become addicted to external validation through likes and reach.

Outsourcing to AI: Convenience vs. Repercussions As AI becomes more integrated into everyday life, there is a growing trend to outsource essential cognitive tasks to machines. This raises critical questions about which cognitive tasks we should protect and preserve [19, 29]. Take ChatGPT, for example: while it seems harmless, generating ideas and text from prompts, it raises deeper concerns. Humans take pride in their ability to think, reflect, and act deliberately - skills that AI is designed to mimic, not replace. But are we still thinking, reflecting, and acting according to our beliefs? Or are we slowly turning into a society that seeks recognition and reward without the willingness to assume the risks, responsibilities, and effort involved? Why bother thinking when ChatGPT can solve the problem for me? Even if we begin to use this approach for mundane or unimportant tasks - what will stop me from outsourcing more and more and possibly caring less and less?

Consider the change in writing: a year and a half ago we had to write all texts ourselves. Now it is possible to scribble some notes and let ChatGPT do the rest. The output is often so polished that we have to remind ourselves to critically

evaluate its semantic meaning. Over time, this tendency toward efficiency over thoughtfulness may threaten to undermine our capacity for deep understanding and genuine care.

The same is true for reviewing and critiquing papers, publications, and dissertations. As ChatGPT improves the quality of writing, we must resist the urge to skim and instead truly question the validity of the content. We need to actively remind ourselves to engage in "critical thinking mode" to maintain awareness and skepticism, as it takes more energy than just nodding along when the text is easy on the eyes.

While some experts and scientists are disappointed with the results of ChatGPT, noting that it failed to capture their intended message, the concern is greater for those who are already disengaged from their work or for a new generation accustomed to instant gratification. If you have a clear idea of what you want to say, it is difficult for the AI to match it. But the vaguer the idea of what you are trying to say, the better the generated text seems to be. So, who will continue to think deeply? Who will continue to use writing as a tool to deepen their understanding and strengthen their arguments?

Navigating social pressures and implications In many industries, the adoption of AI is becoming a necessity rather than a choice, forcing individuals and companies to embrace these technologies even if they are not fully prepared. As we continue to integrate AI into our societies, it is imperative to balance its benefits with its potential risks.

The critical challenge is to ensure that the development of AI is guided by a comprehensive understanding of its social implications and governed by robust ethical frameworks that prioritize human well-being. This requires a collective effort to educate, regulate, and monitor AI applications to prevent harm. This way, we can ensure that AI advances contribute positively to society, rather than exacerbating existing disparities and creating new forms of inequality.

4.3 Politics and Regulation

In the political realm, the call to action for AI is framed by the need for robust regulation - we must create a regulatory framework that is both effective and adaptive. Currently, our regulatory systems struggle to keep up with the pace of the rapid advances in AI. The fundamental nature of AI development, driven by economic interests prioritizing speed, often outpaces the slower, more deliberate pace of policymaking. Moreover, those developing AI technologies tend to be more dynamic and action-oriented than the regulators tasked with overseeing them. This dynamic creates a regulatory environment that is reactive rather than proactive, with legislation constantly lagging behind technological achievements.

The Need for Effective Governance AI is advancing faster than our understanding of both the technology itself and its broader implications. This mismatch is exacerbated in a world defined by volatility, uncertainty, complexity,

and ambiguity (VUCA) [5], where the consequences of AI are interconnected and influenced by numerous factors that are not always apparent. In this context, predicting and mitigating the medium- to long-term impacts of AI becomes particularly challenging, making the need for effective governance even more critical.

We place significant hope and responsibility on regulators to establish a framework that keeps us safe and on track - a responsibility that regulators cannot and should not bear alone. Fortunately, we are beginning to see progress in this direction.

Current AI Regulation Efforts Governments and international organizations are increasingly recognizing the need to regulate artificial intelligence to ensure its ethical development and deployment. Notable efforts include:

- European Union (EU) AI Act: The EU has proposed the Artificial Intelligence Act, aiming to create a comprehensive regulatory framework for AI. This act classifies AI systems into risk categories (unacceptable, high, limited, and minimal) and imposes strict requirements on high-risk applications, including transparency, accountability, and human oversight.
- United States: The U.S. has introduced various guidelines and frameworks, such as the National AI Initiative Act and the Algorithmic Accountability Act, which focus on promoting innovation while addressing ethical concerns. The National Institute of Standards and Technology has developed the AI Risk Management Framework for trustworthy AI to guide organizations in developing and using AI responsibly.
- China: China has implemented the New Generation Artificial Intelligence Development Plan, emphasizing AI leadership and ethical considerations. The country has also introduced regulations requiring transparency and accountability in AI systems, particularly in sectors like finance and healthcare.
- United Nations (UN): The UN has been advocating for global AI ethics standards through initiatives like UNESCO's Recommendation on the Ethics of Artificial Intelligence. This recommendation outlines principles for ensuring that AI respects human rights, diversity, and privacy.

These regulatory efforts aim to balance innovation with ethical considerations, ensuring AI technologies are developed and used in ways that benefit society while mitigating potential risks.

We should not forget, however, that governments and regulators always lag behind as technology generally needs to be introduced first to get regulated. This comes with issues as certain technologies cannot be successfully restricted once released to the public without prior safeguards and restrictions. In addition, regulators often do not have the technical expertise to truly evaluate the risks and potential consequences of technologies. You can watch endless hours of congressional hearings showcasing the often-present naivety. AI, being the black box that it is, definitely does not make things easier.

Thus, we need new frameworks that balance innovation with responsibility, ensuring that AI development benefits society as a whole. We need to find new

ways to ensure that those driving innovation feel accountable and are held accountable for their actions and their consequences, especially when dealing with such invasive and radical capabilities. It cannot be sufficient to argue that it is up to the market and regulators to set and control the conditions for safe and secure technologies. We must ensure that responsibility and accountability are embedded from the outset.

5 Conclusions and Perspectives

Today's AI is characterized by its vast potential and myriad interpretations. At its core, AI raises fundamental questions about what we can achieve with technology, challenging our understanding of capabilities and ethical boundaries. This discussion delves into the technological possibilities of AI, moving beyond basic functionalities to explore its unexpected and sometimes startling emergent properties and unanticipated problems.

What exactly is AI? Definitions vary widely, shaped by the different perceptions of experts in different fields. Each perspective, whether from academic leaders or industry innovators, adds layers to our understanding of what AI can and cannot do. This diversity of viewpoints enriches the discourse, but also complicates the narrative, underscoring the need for a nuanced exploration of AI's capabilities to derive a more holistic view and lay the foundation for a shared understanding.

The inner workings of AI produce such remarkable results yet remain somewhat mysterious. Although the mechanisms behind their success are not fully understood, the combination of massive computing power and massive data has led to impressive results. With sufficient investment and a few AI experts, big tech companies can access the know-how, computing power and data needed to effectively train these models.

AIs learn to categorize data based on their training sets. An integrated and automated learning loop refines this categorization until the AI produces meaningful results. Experts often compare this process to how children learn to walk and talk: through observation, trial and error, and feedback. This feedback strengthens the successful connections and weakens those that do not work, creating a network of connected nodes of varying sizes based on their statistical importance within the training data set. This also explains and underlines the importance of sufficiently large and meaningful data sets.

However, because AI outputs are not based on predefined rules or formalized knowledge, it is unclear what the model has observed and learned. This opacity makes the results volatile; they can be both remarkably innovative and frustratingly inadequate [19]. Unlike specialized tools like AlphaGo and AlphaFold, generic AI applications often exhibit a wider quality gap in terms of accuracy and utility. This gap underscores the importance of users being experts or at least having a basic understanding of their tasks in order to effectively assess the appropriateness and accuracy of AI-generated results.

While we may be clear about the knowledge, we want to encode in the training set, the models may recognize and learn entirely different patterns that are also present but unnoticed or seem irrelevant to us. There are numerous cases where image recognition has failed because AIs learned different patterns than intended. For example, an AI trained to distinguish between men and women repeatedly failed to identify black women as women. It was later discovered that the model was not differentiating based on gender, but on the presence of makeup. Since black women were less likely to wear makeup in the dataset, the model misclassified them as "makeup-free" and therefore "male" [32].

This situation is reflected in the interviews which present a spectrum of opinions from interdisciplinary experts on the future of AI, from optimistic to cautious and skeptical. Some emphasize the limitless potential of AI, while others stress the need for ethical considerations and human oversight. The consensus underscores the importance of building trust through transparency, accountability, and education. Ensuring the ethical use of AI includes addressing bias, regulatory oversight, and upholding human-centred values.

The discussion around the societal implications of AI extends to its impact on human interactions, cognitive tasks, and social dynamics. There are concerns that reliance on AI could lead to cognitive atrophy and dehumanization. Balancing the benefits of AI with its potential risks requires comprehensive education, regulation, and ethical frameworks.

There is also consensus that effective regulation is critical to managing the rapid development of AI. Governments and international organizations are beginning to recognize the need for robust frameworks to ensure the ethical use of AI. However, regulators often lag behind technological advances, requiring a proactive and adaptive approach.

All experts agree that AI offers tremendous potential, but that it also raises fundamental questions about the technology's capabilities and ethical boundaries. It was repeatedly emphasized that interdisciplinary collaboration is required to guarantee a responsible development of AI that embodies ethical considerations, and effective regulation to balance technological progress with societal values.

In summary, the rapid advances in AI have fundamentally challenged our notions of creativity, reason, and consciousness. AI's ability to generate content quickly and efficiently is forcing us to rethink what makes us uniquely human. In fact, we believe that it is time to change perspective and to regard advanced AIs not as tools that we can control but as partners that help us to deal with today's challenges.

This change of perspective has a number of consequences. It shifts the focus of explainable AI from a "how" perspective to a "why" perspective. Instead of asking how an AI arrived at a decision, we should ask why. And, in fact, AIs are very good in this kind of posthoc rationalization [26]. On the other hand, pinpointing the exact "how" is often infeasible. Even if we knew the "how", this technical knowledge would hardly be helpful for, e.g., increasing trust in a proposed decision.

Viewing AIs as partners naturally comes with questions about their limitations concerning versatility, adaptability, creativity, logical reasoning, consciousness, self-awareness, and autonomy. We have seen signs of almost all these properties in recent LLMs. They are capable of many tasks [16, 14], adapt to situations, e.g., language of the prompts, and request of the output format, create texts and pictures, perform elementary arithmetic and some logical reasoning (though not always correctly), are even able to answer questions about themselves and to draw self-portraits [43, 36, 44], and can even autonomously play entire Chess or Go games beyond human capabilities.

For some interviewees this already justifies classifying them as AGIs, whereas others speak of advanced forms of automation and regard AGIs as an unrealistic hypothetical concept. Thus, we are far from having a common, interdisciplinary understanding even of the basic vocabulary.

Alan Turing tried already in the 1950s to make the notion of artificial intelligence tangible by proposing what is now known as a Turing Test [47][4]. The Turing Test focuses on the conversational behaviour of AIs. Later, a number of enhancements were proposed to test creativity (Lovelace Test [7]), self-awareness (Metzinger Test [30]), or deep understanding and integration of knowledge addressing, e.g., moral dilemmas or hypothetical scenarios (Marcus Test [28]).

As mentioned, there is no agreement whether and which of these tests are passed by advanced LLMs. Establishing objective criteria for when such a test is successful would therefore be a major step towards a common interdisciplinary understanding. Moreover, none of the known tests address autonomy or the ability to reflect and self-correct which is therefore another important future challenge.

References

1. Ahrendt, W., Steffen, B.: Let's Talk AI with Wolfgang Ahrendt. In: Barbara Steffen, Edward A. Lee, Bernhard Steffen (eds.) Let's Talk AI: Interdisciplinarity Is a Must, LNCS, vol. 15000, p. (this volume). Springer Nature (2024)
2. Bartlett, S.: Emergency Episode: Ex-Google officer finally speaks out on the the dangers of AI! YouTube video (June 2023), `\url{https://www.youtube.com/watch?v=bk-nQ7HF6k4}`, the Diary of a CEO
3. Baum, K., Steffen, B.: Let's Talk AI with Kevin Baum. In: Barbara Steffen, Edward A. Lee, Bernhard Steffen (eds.) Let's Talk AI: Interdisciplinarity Is a Must, LNCS, vol. 15000, p. (this volume). Springer Nature (2024)
4. Bengio, Y.: Reasoning Through Arguments Against Taking AI Safety Seriously. `https://yoshuabengio.org/2024/07/09/reasoning-through-arguments-against-taking-ai-safety-seriously/` (July 9 2024), accessed: 2024-08-19
5. Bennett, N., Lemoine, G.J.: What a Difference a Word Makes: Understanding Threats to Performance in a VUCA World. Business Horizons **57**(3), 311–317 (2014). https://doi.org/10.1016/j.bushor.2014.01.001
6. Borges, G., Steffen, B.: Let's Talk AI with Georg Borges. In: Barbara Steffen, Edward A. Lee, Bernhard Steffen (eds.) Let's Talk AI: Interdisciplinarity Is a Must, LNCS, vol. 15000, p. (this volume). Springer Nature (2024)

[4] His original Imitation Game was in fact a bit more sophisticated.

7. Bringsjord, S., Bello, P., Ferrucci, D.: Creativity, the Turing test, and the (better) Lovelace test. The Turing test: the elusive standard of artificial intelligence pp. 215–239 (2003)
8. Coombe, D., Curtis, V., Orlowski, J.: The Social Dilemma. Netflix (2020), directed by Jeff Orlowski. Produced by Exposure Labs.
9. Dale, V., McEwan, M., Bohan, J.: Early adopters versus the majority: Characteristics and implications for academic development and institutional change. Journal of Perspectives in Applied Academic Practice **9**(2), 54–67 (2021)
10. Enkel, E., Steffen, B.: Let's Talk AI with Ellen Enkel. In: Barbara Steffen, Edward A. Lee, Bernhard Steffen (eds.) Let's Talk AI: Interdisciplinarity Is a Must, LNCS, vol. 15000, p. (this volume). Springer Nature (2024)
11. Fey, M., Steffen, B.: Let's Talk AI with Matthias Fey. In: Barbara Steffen, Edward A. Lee, Bernhard Steffen (eds.) Let's Talk AI: Interdisciplinarity Is a Must, LNCS, vol. 15000, p. (this volume). Springer Nature (2024)
12. Grallet, G., Pons, H.: Yuval Noah Harari (Sapiens) versus Yann LeCun (Meta) on Artificial Intelligence. Le Point (May 11 2023), `https://www.lepoint.fr/sciences-nature/yuval-harari-sapiens-versus-yann-le-cun-meta-on-artificial-intelligence-11-05-2023-2519782\_1924.php`, accessed: 2024-08-19
13. Güntürkün, O., Steffen, B.: Let's Talk AI with Onur Güntürkün. In: Barbara Steffen, Edward A. Lee, Bernhard Steffen (eds.) Let's Talk AI: Interdisciplinarity Is a Must, LNCS, vol. 15000, p. (this volume). Springer Nature (2024)
14. Harari, Y.B..Y.N.: Artificial Intelligence, Democracy, & the Future of Civilization. YouTube video (May 2023), `\url{https://www.youtube.com/watch?v=TKopbyIPo6Y}`, recorded on May 24 2023, as part of the C2 Montreal conference, presented by Mila - Quebec Artificial Intelligence Institute
15. Harris, T.: Beyond the AI dilemma. YouTube video (June 2023), `\url{https://www.youtube.com/watch?v=e5dQ5zEuE9Q\&t=1443s}`, cogX Festival 2023
16. Harris, T., Raskin, A.: The AI dilemma. Online Video (March 2023), `https://www.youtube.com/watch?v=xoVJKj8lcNQ`, accessed: 2024-08-10
17. Helfer, T., Steffen, B.: Let's Talk AI with Thorsten Helfer. In: Barbara Steffen, Edward A. Lee, Bernhard Steffen (eds.) Let's Talk AI: Interdisciplinarity Is a Must, LNCS, vol. 15000, p. (this volume). Springer Nature (2024)
18. Hermanns, H., Steffen, B.: Let's Talk AI with Holger Hermanns. In: Barbara Steffen, Edward A. Lee, Bernhard Steffen (eds.) Let's Talk AI: Interdisciplinarity Is a Must, LNCS, vol. 15000, p. (this volume). Springer Nature (2024)
19. Hernández-Orallo, J., Steffen, B.: Let's Talk AI with José Hernández-Orallo. In: Barbara Steffen, Edward A. Lee, Bernhard Steffen (eds.) Let's Talk AI: Interdisciplinarity Is a Must, LNCS, vol. 15000, p. (this volume). Springer Nature (2024)
20. Hinchey, M., Steffen, B.: Let's Talk AI with Mike Hinchey. In: Barbara Steffen, Edward A. Lee, Bernhard Steffen (eds.) Let's Talk AI: Interdisciplinarity Is a Must, LNCS, vol. 15000, p. (this volume). Springer Nature (2024)
21. Howar, F., Steffen, B.: Let's Talk AI with Falk Howar. In: Barbara Steffen, Edward A. Lee, Bernhard Steffen (eds.) Let's Talk AI: Interdisciplinarity Is a Must, LNCS, vol. 15000, p. (this volume). Springer Nature (2024)
22. Johnson, T., Steffen, B.: Let's Talk AI with Taylor Johnson. In: Barbara Steffen, Edward A. Lee, Bernhard Steffen (eds.) Let's Talk AI: Interdisciplinarity Is a Must, LNCS, vol. 15000, p. (this volume). Springer Nature (2024)
23. Katoen, J.P., Steffen, B.: Let's Talk AI with Joost-Pieter Katoen. In: Barbara Steffen, Edward A. Lee, Bernhard Steffen (eds.) Let's Talk AI: Interdisciplinarity Is a Must, LNCS, vol. 15000, p. (this volume). Springer Nature (2024)

24. Krämer, N., Steffen, B.: Let's Talk AI with Nicole Krämer. In: Barbara Steffen, Edward A. Lee, Bernhard Steffen (eds.) Let's Talk AI: Interdisciplinarity Is a Must, LNCS, vol. 15000, p. (this volume). Springer Nature (2024)
25. Langer, M., Steffen, B.: Let's Talk AI with Markus Langer. In: Barbara Steffen, Edward A. Lee, Bernhard Steffen (eds.) Let's Talk AI: Interdisciplinarity Is a Must, LNCS, vol. 15000, p. (this volume). Springer Nature (2024)
26. Lee, E., Steffen, B.: Let's Talk AI with Edward Lee. In: Barbara Steffen, Edward A. Lee, Bernhard Steffen (eds.) Let's Talk AI: Interdisciplinarity Is a Must, LNCS, vol. 15000, p. (this volume). Springer Nature (2024)
27. Leucker, M., Steffen, B.: Let's Talk AI with Martin Leucker. In: Barbara Steffen, Edward A. Lee, Bernhard Steffen (eds.) Let's Talk AI: Interdisciplinarity Is a Must, LNCS, vol. 15000, p. (this volume). Springer Nature (2024)
28. Marcus, G.: The Search for a New Test of Artificial Intelligence. Scientific American (March 1 2017), https://www.scientificamerican.com/article/the-search-for-a-new-test-of-artificial-intelligence/, accessed: 2024-08-19
29. Margaria, T., Steffen, B.: Let's Talk AI with Tiziana Margaria. In: Barbara Steffen, Edward A. Lee, Bernhard Steffen (eds.) Let's Talk AI: Interdisciplinarity Is a Must, LNCS, vol. 15000, p. (this volume). Springer Nature (2024)
30. Metzinger, P.D.T.: Postbiotisches Bewusstsein: Wie man ein künstliches Subjekt baut - und warum wir es nicht tun sollten. https://www.hnf.de/veranstaltungen/events/paderborner-podium/computer-gehirn-und-bewusstsein/metzinger.html, accessed: 2024-08-19
31. Neider, D., Steffen, B.: Let's Talk AI with Daniel Neider. In: Barbara Steffen, Edward A. Lee, Bernhard Steffen (eds.) Let's Talk AI: Interdisciplinarity Is a Must, LNCS, vol. 15000, p. (this volume). Springer Nature (2024)
32. Noah, T., Hoffmann, R., Finger, A.: Trevor Noah on the Future of Entertainment and AI. Interview on the Possible Podcast. https://www.youtube.com/watch?v=4ZD-ZIS2CfU (2023), accessed: 2024-08-19
33. Olsberg, K.: Virtua: KI – Kontrolle ist Illusion. Aufbau Taschenbuch, Berlin, Germany (2023)
34. Pelley, S.: "Godfather of Artificial Intelligence" Geoffrey Hinton on the Promise, Risks of Advanced AI. CBS News, 60 Minutes - Newsmakers (June 16 2024), https://www.cbsnews.com/news/geoffrey-hinton-ai-dangers-60-minutes-transcript/, accessed: 2024-08-19
35. Perrigo, B.: Yann LeCun On How An Open Source Approach Could Shape AI. Time (February 7 2024), https://time.com/6691705/time100-impact-awards-yann-lecun/, accessed: 2024-08-19
36. Pi, Steffen, B.: Let's Talk AI with Pi. In: Barbara Steffen, Edward A. Lee, Bernhard Steffen (eds.) Let's Talk AI: Interdisciplinarity Is a Must, LNCS, vol. 15000, p. (this volume). Springer Nature (2024)
37. Rehof, J., Steffen, B.: Let's Talk AI with Jakob Rehof. In: Barbara Steffen, Edward A. Lee, Bernhard Steffen (eds.) Let's Talk AI: Interdisciplinarity Is a Must, LNCS, vol. 15000, p. (this volume). Springer Nature (2024)
38. Ries, E.: The Lean Startup. https://www.youtube.com/watch?v=fEvKo90qBns (2011), youTube video, Accessed: 2024-08-19
39. Schieferdecker, I., Steffen, B.: Let's Talk AI with Ina Schieferdecker. In: Barbara Steffen, Edward A. Lee, Bernhard Steffen (eds.) Let's Talk AI: Interdisciplinarity Is a Must, LNCS, vol. 15000, p. (this volume). Springer Nature (2024)
40. Schmidt, E., Steffen, B.: Let's Talk AI with Eva Schmidt. In: Barbara Steffen, Edward A. Lee, Bernhard Steffen (eds.) Let's Talk AI: Interdisciplinarity Is a Must, LNCS, vol. 15000, p. (this volume). Springer Nature (2024)

41. Sesing-Wagenpfeil, A., Steffen, B.: Let's Talk AI with Andreas Sesing-Wagenpfeil. In: Barbara Steffen, Edward A. Lee, Bernhard Steffen (eds.) Let's Talk AI: Interdisciplinarity Is a Must, LNCS, vol. 15000, p. (this volume). Springer Nature (2024)
42. Speith, T., Steffen, B.: Let's Talk AI with Timo Speith. In: Barbara Steffen, Edward A. Lee, Bernhard Steffen (eds.) Let's Talk AI: Interdisciplinarity Is a Must, LNCS, vol. 15000, p. (this volume). Springer Nature (2024)
43. Steffen, B.: How Hot is the Water. In: Barbara Steffen, Edward A. Lee, Bernhard Steffen (eds.) Let's Talk AI: Interdisciplinarity Is a Must, LNCS, vol. 15000, p. (this volume). Springer Nature (2024)
44. Steffen, B.: Let's Talk AI with ChatGPT. In: Barbara Steffen, Edward A. Lee, Bernhard Steffen (eds.) Let's Talk AI: Interdisciplinarity Is a Must, LNCS, vol. 15000, p. (this volume). Springer Nature (2024)
45. Steffen, B., Steffen, B.: Let's Talk AI with Bernhard Steffen. In: Barbara Steffen, Edward A. Lee, Bernhard Steffen (eds.) Let's Talk AI: Interdisciplinarity Is a Must, LNCS, vol. 15000, p. (this volume). Springer Nature (2024)
46. Suleyman, M.: The Coming Wave: Technology, Power, and the Twenty-First Century's Greatest Dilemma. Crown Publishing Group, New York, NY (2023)
47. Turing, A.M.: Computing Machinery and Intelligence. Mind **59**(236), 433–460 (1950). https://doi.org/10.1093/mind/LIX.236.433
48. Vardi, M., Steffen, B.: Let's Talk AI with Moshe Vardi. In: Barbara Steffen, Edward A. Lee, Bernhard Steffen (eds.) Let's Talk AI: Interdisciplinarity Is a Must, LNCS, vol. 15000, p. (this volume). Springer Nature (2024)
49. Weizenbaum Institute: Weizenbaum institute for the networked society (2024), `https://www.weizenbaum-institut.de`, accessed: 2024-08-10
50. von Wendt, K., Steffen, B.: Let's Talk AI with Karl von Wendt. In: Barbara Steffen, Edward A. Lee, Bernhard Steffen (eds.) Let's Talk AI: Interdisciplinarity Is a Must, LNCS, vol. 15000, p. (this volume). Springer Nature (2024)
51. Wien, T.: Digital Humanism Initiative. `https://dighum.ec.tuwien.ac.at/`, accessed: 2024-08-19
52. Wirsing, M., Steffen, B.: Let's Talk AI with Martin Wirsing. In: Barbara Steffen, Edward A. Lee, Bernhard Steffen (eds.) Let's Talk AI: Interdisciplinarity Is a Must, LNCS, vol. 15000, p. (this volume). Springer Nature (2024)