

Enhancing Trust and Transparency in AI-Generated Visual Content: A Comprehensive Framework for Authentic Media Production

Sarah Ameen Hezam
Student ID: 222000318

August 2026

Abstract

The rapid proliferation of generative artificial intelligence has fundamentally transformed visual media production, enabling the creation of highly realistic images and videos through tools like DALL-E, Midjourney, and Sora. However, this technological advancement has created a significant challenge: audiences increasingly struggle to distinguish between real and AI-generated content, leading to what scholars describe as an "authenticity crisis" that undermines trust in media. This research addresses the critical gap between the technical capabilities of generative AI and the societal need for transparent, trustworthy visual content. Building upon a comprehensive literature review conducted in previous assignments, this study employs a mixed-methods approach that combines a systematic literature review, qualitative interviews with content creators and media professionals, and quantitative surveys of audience perceptions. The principal contribution of this work is the development of a novel framework designed to enhance trust and transparency in AI-generated visual content. This framework, informed by interdisciplinary insights from computer science, media studies, and communication theory, introduces pragmatic guidelines for content labeling, transparency mechanisms, and audience engagement. Additionally, the research provides practical recommendations for content creators, platform designers, and policymakers, contributing to the growing body of knowledge on ethical AI and fostering a more transparent and trustworthy digital media landscape.

Keywords: Generative AI, Visual Content, Trust, Transparency, Authenticity Crisis, Media Production, Audience Perception, Ethical AI Framework

Contents

1	Introduction	4
2	Background	4
2.1	The Evolution of Generative AI in Visual Content	4
2.2	Trust and Transparency Issues in AI-Generated Visual Content	5
2.3	The Significance of Trust and Transparency	5
2.4	Challenges to Enhancing Trust in AI-Generated Visual Content	6
3	Problem Statement	6
4	Research Objectives	7
4.1	Objective 1: To comprehensively analyze the existing challenges in trust and transparency of AI-generated visual content	7
4.2	Objective 2: To evaluate existing solutions, including content labeling, watermarking, and Explainable AI (XAI) frameworks	7
4.3	Objective 3: To propose a comprehensive framework for enhancing trust in AI-generated visual content	8
4.4	Objective 4: To validate the proposed framework in diverse visual content domains	8
4.5	Objective 5: To provide practical recommendations for the implementation of the proposed framework	9
5	Research Scope	9
6	Literature Review	10
6.1	The Importance of Trust and Transparency in AI-Generated Visual Content	10
6.2	Current Challenges and Limitations	10
6.3	Existing Solutions: XAI and Trustworthy AI Frameworks	11
6.4	Gaps in the Literature	11
6.5	Relevance to This Research	12
7	Research Methodology	12
7.1	Phase 1: Literature Review (Completed)	12
7.2	Phase 2: Qualitative Research	13
7.3	Phase 3: Quantitative Research	13
7.4	Phase 4: Framework Development and Validation	13
8	Ethical Considerations	14
9	Limitations	14

10 Expected Outcomes	15
11 Timeline (Gantt Chart)	16
References	18

1 Introduction

In recent years, generative artificial intelligence has fundamentally reshaped the landscape of visual media production. Tools such as DALL-E, Midjourney, and Sora can now produce images and videos that are often indistinguishable from real photographs or professionally shot footage (ERTAN & SAHIN CEKEN, 2023). This technological leap has democratized content creation, enabling anyone with a computer to generate high-quality visuals. However, as I observed when reviewing the literature for my previous assignments, this progress has also introduced a serious problem: audiences are losing the ability to tell what is real and what is machine-made.

The significance of trust in visual media cannot be overstated. In journalism, advertising, education, and even personal relationships, the authenticity of visual content is foundational to credibility and decision-making. When this authenticity is compromised, the consequences are far-reaching. Reddy (2022) and Zhang and Lu (2021) have articulated the challenges posed by opaque AI systems, emphasizing the urgent need for frameworks that enhance transparency and trustworthiness.

This research is motivated by the critical gap between the current capabilities of generative AI and the societal need for transparent, ethical, and trustworthy visual content. While significant advancements have been made in the technical capabilities of generative AI models, there remains a gap in understanding how to effectively communicate the AI-generated nature of visual content to audiences without diminishing its creative and practical value (Tran, 2025; Duong & Vo, 2025). Existing approaches to content labeling and transparency are often fragmented, platform-dependent, and lack empirical validation across diverse contexts (Calli & Calli, 2025).

Therefore, this research aims to address this gap by proposing a comprehensive framework that enhances trust and transparency in AI-generated visual content across multiple dimensions: technical, ethical, and practical. This framework seeks to align with the needs of various stakeholders, including content creators, platform designers, policymakers, and audiences, fostering a more ethical and transparent media landscape.

2 Background

2.1 The Evolution of Generative AI in Visual Content

Generative AI has transitioned from a niche research area to a mainstream tool for content creation. The development of Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and more recently, diffusion models, has enabled the synthesis of high-fidelity images and videos from textual descriptions (Zhang et al., 2025). Tools like DALL-E, Midjourney, and Stable Diffusion have made these capabilities accessible to the

public, leading to an explosion of AI-generated content on social media, advertising, and entertainment platforms.

2.2 Trust and Transparency Issues in AI-Generated Visual Content

Trust in media is built on the expectation that what we see accurately represents reality. However, the opacity of generative AI models poses significant challenges to this expectation. The "black box" nature of these models means that even their creators may not fully understand how specific outputs are generated. This lack of transparency is problematic in any context but is particularly critical in visual media, where images and videos carry significant emotional and informational weight.

The literature reveals several key issues:

- **Authenticity Crisis:** Audiences struggle to distinguish between real and AI-generated images, leading to a general erosion of trust in visual media (ERTAN & SAHIN CEKEN, 2023).
- **Emotional Disconnect:** Studies show that AI-generated content is perceived as efficient but lacks the emotional depth of human-created content, resulting in lower engagement (Tran, 2025).
- **Generational Differences:** Generation Z users value emotional authenticity and narrative coherence, and they express discomfort when authorship is undisclosed (Duong & Vo, 2025).
- **Platform-Specific Responses:** Public discourse on platforms like YouTube reveals emotional and cognitive anxiety, including the "uncanny valley" effect and epistemic uncertainty about what is real (Calli & Calli, 2025).

2.3 The Significance of Trust and Transparency

Trust is the foundation of any successful media ecosystem. For AI-generated visual content to reach its full potential, it must be adopted and accepted by audiences, and this acceptance is heavily contingent upon trust. Trustworthy AI-generated content is essential not only for individual user comfort and safety but also for the societal adoption of generative technologies. Furthermore, as regulatory bodies around the world begin to scrutinize AI more closely, the ability to demonstrate the transparency and ethical grounding of AI systems becomes a legal and ethical imperative.

2.4 Challenges to Enhancing Trust in AI-Generated Visual Content

Enhancing trust in AI-generated visual content involves addressing both technical and ethical challenges:

- **Technical Challenges:** Developing methods to make generative AI systems more interpretable and explainable remains a significant research focus. This includes improving content labeling, watermarking, and metadata embedding.
- **Ethical Challenges:** Ensuring that AI systems make decisions and produce content in a manner that aligns with human values and ethics is paramount. This includes addressing issues of bias, fairness, and the potential for misuse.
- **Practical Challenges:** Implementing transparency mechanisms consistently across different platforms and contexts, and ensuring that these mechanisms are actually understood and trusted by audiences.

3 Problem Statement

In the rapidly evolving landscape of generative AI, the lack of transparency and trust in AI-generated visual content hinders its widespread adoption and positive societal impact. The "black box" nature of many advanced generative AI models, as highlighted by ERTAN & SAHIN CEKEN (2023) and Zhang et al. (2025), poses significant challenges in ensuring accountability, ethical usage, and understanding of the content creation process. Existing solutions, such as content labeling systems and watermarking techniques, have faced criticism for their fragmentation and lack of empirical validation across diverse real-world contexts (Ghassemi et al., 2022).

The significance of this problem is underscored by its impact on critical domains like journalism, education, advertising, and entertainment, where trust and transparency are paramount. As AI increasingly augments or replaces traditional content creation, the inability to explain or authenticate AI-generated media erodes audience trust (Reddy, 2022). The problem domain extends beyond technical issues, encompassing management complexities (Berente et al., 2021), the potential of AI to contribute to sustainable development (Goralski & Tan, 2020), and the broader societal implications outlined in the One Hundred Year Study on Artificial Intelligence (Stone et al., 2022).

From my perspective, this gap is what motivates my research. I believe that trust in AI-generated content cannot be achieved through technical fixes alone. We also need to understand how different audiences—especially younger generations who grew up with social media—perceive transparency and authenticity. Therefore, my research aims to address the following problem:

”How can we develop a comprehensive and practically implementable framework that enhances trust and transparency in AI-generated visual content, ensuring audience awareness and authenticity while preserving the creative and practical benefits of generative AI?”

4 Research Objectives

To address the problem stated above, this research pursues the following five objectives, structured to move logically from understanding the problem to proposing and validating a practical solution:

4.1 Objective 1: To comprehensively analyze the existing challenges in trust and transparency of AI-generated visual content

This objective focuses on building a solid foundation by identifying what we already know about trust issues in AI-generated media. Specifically, I aim to:

- Identify and categorize the key challenges audiences face when encountering AI-generated images and videos, based on existing literature.
- Examine how issues of authenticity, bias, and platform-specific transparency mechanisms affect user trust.
- Synthesize findings from different audience groups (e.g., Gen Z, general social media users, professional content creators) to identify common patterns and differences.

Justification: Without a clear understanding of the current challenges, any proposed solution would be guesswork. This objective ensures my research is grounded in real issues rather than assumptions.

4.2 Objective 2: To evaluate existing solutions, including content labeling, watermarking, and Explainable AI (XAI) frameworks

Once the challenges are mapped, the next step is to look at what has already been tried. This objective involves:

- Critically assessing the strengths and weaknesses of current transparency mechanisms such as visible labels, invisible watermarks, and metadata embedding.

- Evaluating how well Explainable AI (XAI) methods—originally designed for general AI systems—can be adapted to generative visual content.
- Exploring alternative validation approaches proposed by scholars such as Ghassemi et al. (2022), who argue that current XAI methods have significant imperfections.

Justification: This ensures that my proposed framework does not reinvent the wheel but instead builds upon and improves existing approaches.

4.3 Objective 3: To propose a comprehensive framework for enhancing trust in AI-generated visual content

This is the core contribution of my research. Based on the analysis from Objectives 1 and 2, I will develop a framework that:

- Integrates technical transparency mechanisms (labels, watermarks, metadata) with human-centered design principles that consider how audiences actually perceive and respond to transparency cues.
- Addresses the specific needs of different platforms (social media, advertising, cinematic production) and audience groups.
- Aligns with ethical considerations, including informed consent, user autonomy, and fairness.

Justification: This objective transforms the research from a purely analytical exercise into a practical contribution that can be used by content creators, platform designers, and policymakers.

4.4 Objective 4: To validate the proposed framework in diverse visual content domains

A framework is only useful if it actually works in the real world. Therefore, this objective involves:

- Testing the framework across different types of visual content platforms, including social media (TikTok, YouTube, Instagram), digital advertising, and cinematic production.
- Gathering feedback from practitioners, content creators, and media professionals to refine the framework based on practical experience.
- Assessing the generalizability of the framework and identifying any domain-specific adjustments that may be needed.

Justification: This ensures that my proposed framework is not just theoretical but has practical value and can be adapted to different contexts.

4.5 Objective 5: To provide practical recommendations for the implementation of the proposed framework

The final objective ensures that my research has a real-world impact. I will:

- Offer actionable guidance for organizations, content creators, platform developers, and policymakers on how to integrate the framework into their existing workflows.
- Provide a roadmap for implementation that considers resource constraints, technical requirements, and potential challenges.
- Contribute to the ongoing academic and industry discourse on responsible and trustworthy AI development and deployment.

Justification: Without practical recommendations, even the best framework may remain unused. This objective bridges the gap between academic research and real-world practice.

5 Research Scope

This research focuses on the following areas to ensure a manageable and impactful study:

- **Analysis of Existing Challenges:** Concentrating on identifying and understanding the key issues that currently hinder trust and transparency in AI-generated visual content, with a particular emphasis on social media platforms (TikTok, YouTube, Instagram), digital advertising, and cinematic production.
- **Evaluation of Current Solutions:** Exploring existing solutions, including but not limited to, content labeling systems, watermarking techniques, metadata embedding, and Explainable AI (XAI) frameworks. This evaluation seeks to assess the efficacy and limitations of these solutions in addressing the identified challenges specifically within the visual content domain.
- **Development of a New Framework:** Proposing a novel framework that integrates insights from the analysis and evaluations. This framework aims to offer practical and ethical guidelines to enhance trustworthiness in AI-generated visual content, considering both technical and human-centered factors.
- **Validation in Diverse Contexts:** Applying the proposed framework to varied visual content domains (social media, advertising, cinematic production) to assess its generalizability and effectiveness. This step includes seeking feedback from a range

of stakeholders, including content creators, media professionals, platform designers, and audiences.

6 Literature Review

The literature review is organized into several key themes that provide the foundation for this research.

6.1 The Importance of Trust and Transparency in AI-Generated Visual Content

The rapid advancement of generative AI has transformed visual media production, but it has also created a significant authenticity crisis (ERTAN & SAHIN CEKEN, 2023). Audiences now struggle to distinguish between real and AI-generated images and videos, which directly undermines trust in media. Key findings from recent audience studies highlight several patterns:

- AI-generated content is perceived as efficient but lacks emotional depth compared to human-made content (Tran, 2025).
- Generation Z users value emotional authenticity and narrative coherence, and they express discomfort when authorship is undisclosed (Duong & Vo, 2025).
- Public discourse on platforms like YouTube reveals emotional and cognitive anxiety, including the "uncanny valley" effect and epistemic uncertainty about what is real (Calli & Calli, 2025).

Reddy (2022) and Zhang and Lu (2021) both emphasize that transparency is essential for building trust, whether in clinical AI or media applications. Without transparency, users become skeptical and disengaged.

6.2 Current Challenges and Limitations

Despite progress in AI capabilities, several persistent challenges hinder trust and transparency. Key issues include:

- **Bias and algorithmic accountability**, which raise concerns about fairness and discrimination (Goralski & Tan, 2020).
- **Opacity, autonomy, and inscrutability**, which make it difficult for users to understand or challenge AI decisions (Berente et al., 2021).

- **Difficulty in explaining visual AI decisions**, as images are processed emotionally and quickly, making transparency even more critical compared to text-based systems (Ghassemi et al., 2022).
- **Fragmented transparency mechanisms** that vary across different platforms, leading to inconsistent user experiences (Calli & Calli, 2025).

From my perspective, these challenges are especially critical in visual media because people process images quickly and emotionally. When they see a realistic image, they instinctively trust it—which makes fake content particularly dangerous.

6.3 Existing Solutions: XAI and Trustworthy AI Frameworks

Several solutions have been proposed to address transparency issues, each with distinct strengths and limitations:

Explainable AI (XAI):

- The goal is to make AI decision-making processes understandable to users.
- For example, Reddy (2022) advocates for XAI in clinical settings to uphold informed consent.
- However, Ghassemi et al. (2022) argue that many XAI methods are imperfect and can provide misleading explanations, calling for alternative validation approaches.

Trustworthy AI (TAI):

- Thiebes, Lins, and Sunyaev (2021) propose five foundational principles: beneficence, non-maleficence, autonomy, justice, and explicability.
- While useful, this framework is general and does not specifically address the unique challenges of visual content where emotional and perceptual factors dominate.

Content Labeling and Watermarking:

- These techniques can clearly signal AI-generated material.
- However, adoption is inconsistent across platforms, and users may ignore or distrust labels if not implemented thoughtfully.

6.4 Gaps in the Literature

Based on my review, I identified three main gaps that this research aims to address:

1. **Limited focus on visual content:** Most trust and transparency research addresses general AI systems, not generative visual media specifically.

2. **Lack of a unified framework:** Existing solutions are fragmented and platform-specific. There is no comprehensive model integrating technical transparency with human-centered audience perception.
3. **Insufficient real-world validation:** Most solutions have been tested in controlled environments rather than actual media platforms where users consume content daily.

Stone et al. (2022) emphasize the need for effective governance mechanisms and interdisciplinary approaches, which reinforces my belief that trust in AI-generated content requires more than just technical fixes.

6.5 Relevance to This Research

The literature reviewed confirms that trust and transparency in AI-generated visual content are urgent and underexplored issues. While XAI and TAI frameworks offer useful starting points, they require adaptation to the specific context of visual media. This research directly addresses the identified gaps by focusing specifically on AI-generated visual content, proposing a comprehensive framework that combines technical transparency with audience perception, and validating the framework across diverse real-world platforms.

7 Research Methodology

This research adopts a mixed-methods approach to comprehensively analyze the existing challenges and solutions related to trust and transparency in AI-generated visual content, and to develop and validate a comprehensive framework for enhancing trust in this domain. This methodology is chosen for its strength in providing a holistic understanding of the research problem by combining the depth of qualitative insights with the breadth of quantitative analysis.

7.1 Phase 1: Literature Review (Completed)

The first phase involved an extensive literature review to identify and synthesize existing research on trust, transparency, and authenticity in AI-generated visual content. This review covered scholarly articles, conference papers, industry reports, and relevant guidelines. The purpose was to map out the current landscape of trust in AI-generated media, pinpointing gaps in knowledge and practice that the proposed framework aims to address.

- **Data Sources:** Academic databases (IEEE Xplore, ACM Digital Library, ScienceDirect, Scopus) and professional publications.
- **Selection Criteria:** Publications from the last ten years, focusing on generative AI, visual content, trust, transparency, authenticity, and audience perception.

- **Analysis Technique:** Thematic analysis to identify recurring themes, challenges, and solutions.

7.2 Phase 2: Qualitative Research

The second phase involves qualitative research through semi-structured interviews and case studies. The goal is to gain deeper insights into the experiences and perceptions of stakeholders regarding trust and transparency in AI-generated visual content.

- **Participants:** A purposive sample of content creators, media professionals, filmmakers, and social media users.
- **Data Collection:** Semi-structured interviews conducted remotely, adhering to ethical research standards including informed consent.
- **Analysis:** Qualitative content analysis to extract themes related to trust challenges, perceptions of AI-generated content, and effective strategies for enhancing transparency.

7.3 Phase 3: Quantitative Research

This phase complements the qualitative insights with quantitative data, assessing the prevalence of identified challenges and evaluating the effectiveness of existing solutions among broader audiences.

- **Survey Design:** A survey based on the qualitative phase findings, targeting a broader audience of social media users and consumers of visual content.
- **Sample:** Random sampling of social media users across different age groups and demographics, aiming for statistical significance in the responses.
- **Data Analysis:** Statistical analysis, including descriptive statistics and inferential statistics (e.g., regression analysis, correlation), to understand the distribution of perceptions and the relationship between variables related to trust in AI-generated visual content.

7.4 Phase 4: Framework Development and Validation

Based on insights gathered from the previous phases, a comprehensive framework for enhancing trust in AI-generated visual content will be developed. This framework will be validated through expert reviews and pilot testing in real-world media contexts.

- **Development:** Integration of identified challenges, ethical considerations, and effective transparency strategies into a coherent framework.

- **Validation:** Expert feedback from academics and practitioners in the fields of AI, media studies, and communication, followed by pilot testing in selected social media and advertising contexts.
- **Evaluation Criteria:** Clarity, applicability, comprehensiveness, and effectiveness in enhancing trust.

8 Ethical Considerations

Throughout this research, ethical considerations are paramount:

- **Informed Consent:** All participants in interviews and surveys will be provided with clear information about the research purpose, procedures, risks, and benefits. They will be required to provide explicit consent before participation.
- **Confidentiality and Anonymity:** Participants' identities and data will be protected through anonymization. All data will be stored securely and accessible only to the researcher.
- **Minimizing Harm:** The research will not involve any experimental treatments that could cause physical or psychological harm to participants.
- **Transparency:** The researcher will disclose any potential conflicts of interest and will be transparent about the use of AI tools in the research process, if applicable.
- **Avoiding Plagiarism and Fabrication:** All sources will be properly cited, and data will be reported accurately without fabrication or falsification. Authorship will be determined based on substantial contributions, in line with COPE and ICMJE guidelines.

9 Limitations

This research acknowledges potential limitations, including:

- **Sample Representativeness:** The sample in qualitative research (interviews) will be purposive and may not represent all content creators or media professionals. Efforts will be made to ensure diversity in the sample.
- **Response Bias:** Survey data may be subject to response bias, where participants provide socially desirable answers. Careful questionnaire design and anonymization will help mitigate this.

- **Platform Specificity:** The research may focus on specific platforms (e.g., TikTok, YouTube) which may not fully represent other visual media contexts.
- **Time Constraints:** The research will be conducted within a limited timeframe, which may affect the breadth of data collection.
- **Technical Focus:** While the framework will address technical transparency, it will not delve deeply into the algorithmic details of generative AI models, but rather focus on their practical and perceptual implications.

10 Expected Outcomes

This research aims to produce several significant outcomes:

1. **Comprehensive Analysis of Trust Challenges in AI-Generated Visual Content:** A detailed examination of the current challenges and barriers to trust in AI-generated visual media, highlighting the multifaceted nature of the problem, including technical, perceptual, and ethical dimensions.
2. **Evaluation of Existing Solutions:** An assessment of current methodologies and frameworks (XAI, content labeling, watermarking), identifying their strengths, weaknesses, and gaps in applicability to visual content.
3. **Development of a Novel Framework:** The formulation of an innovative, comprehensive framework for enhancing trust and transparency in AI-generated visual content, providing actionable guidelines for content creators, platform designers, and policymakers.
4. **Practical Recommendations for Stakeholders:** Actionable recommendations tailored to address the specific needs and concerns of developers, users, policymakers, and media professionals, promoting a collaborative effort towards more transparent and ethical AI-generated media.
5. **Contribution to Academic and Industry Discourse:** Enrichment of academic and industry discourse on trust and ethics in AI, stimulating further discussion, research, and innovation in the field.
6. **Foundation for Future Research:** Establishment of a foundation for future research on trust in AI-generated media, identifying new research questions and areas for further investigation.

11 Timeline (Gantt Chart)

The following Gantt chart outlines the proposed timeline for completing this research project over a six-month period. This page is presented in landscape orientation for better readability.

Table 1: Research Project Timeline (Gantt Chart)

Activity	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6
Literature Review (Phase 1)	■					
Qualitative Data Collection (Phase 2)		■				
Qualitative Data Analysis			■			
Quantitative Data Collection (Phase 3)				■		
Quantitative Data Analysis					■	
Framework Development (Phase 4)						■
Framework Validation & Feedback						■
Thesis Writing & Revisions	■	■	■	■	■	■

References

References

- [1] ERTAN, I. F., & SAHIN CEKEN, K. (2023). Generative Artificial Intelligence and Visual Communication: A Systematic Review of Transformations in Image Production, Audience Reception, and Meaning-Making in Digital Media. *Review of Communication Research*, 11, 39.
- [2] Calli, L., & Calli, B. A. (2025). Recoding Reality: A Case Study of YouTube Reactions to Generative AI Videos. *Systems*, 13(10), 925.
- [3] Duong, H. L., & Vo, T. K. O. (2025). How do young users perceive and respond to AI-generated short-form videos? An exploration of Generation Z's perceptions, emotional responses, and trust in AI-created video on social media platforms. *International Journal of Human-Computer Studies*.
- [4] Tran, V. C. (2025). *Bridging Efficiency and Emotion: Evaluating User Engagement with AI-Generated and Human-Generated TikTok Videos Through Media Richness and Uses and Gratifications Theories*. University of Economics Ho Chi Minh City.
- [5] Zhang, R., Yu, B., Min, J., et al. (2025). Generative AI for Film Creation: A Survey of Recent Advances. *arXiv*.
- [6] Reddy, S. (2022). Explainability and artificial intelligence in medicine. *The Lancet Digital Health*, 4(4), e214–e215.
- [7] Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. *Journal of Industrial Information Integration*, 23, 100224.
- [8] Ghassemi, M., Oakden-Rayner, L., & Beam, A. L. (2022). The false hope of current approaches to explainable artificial intelligence in health care. *The Lancet Digital Health*, 4(11), e806–e811.
- [9] Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly*, 45(3), 1433–1450.
- [10] Goralski, M. A., & Tan, T. K. (2020). Artificial intelligence and sustainable development. *The International Journal of Management Education*, 18(1), 100330.
- [11] Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., Hirschberg, J., Kalyanakrishnan, S., Kamar, E., Kraus, S., Leyton-Brown, K., Parkes, D., Press, W., Saxenian, A., Shah, J., Tambe, M., & Teller, A. (2022). Artificial Intelligence and Life in 2030: The One Hundred Year Study on Artificial Intelligence. *arXiv:2211.06318*.

- [12] Thiebes, S., Lins, S., & Sunyaev, A. (2021). Trustworthy artificial intelligence. *Electronic Markets*, 31(3), 447–464.