AI For Humanity

Key Takeaways from the AI For Humanity Track of the 2025 Global Digital Collaboration Conference

Version 1.1

Compiled and edited by Ethan Westfall, GOSIM AI for Humanity Program

July 2025

Sponsored by *Global Open Source Innovation Meetup*

GOSIM

Background

On July 2, 2025, the inaugural AI for Humanity Track at the Global Digital Collaboration (GDC) Conference in Geneva, Switzerland consisted of a series of expert-led presentations, discussions, and panels. The goal of these efforts was to bring together diverse stakeholders from open source and standards organizations, as well as research and academic communities, to begin to shape Responsible AI systems and collaboratively address global challenges.

Note: This report lists speakers' names with their permission; otherwise, Chatham House rules are observed for all participants, per GDC guidelines.

What is GDC?

GDC ("Global Digital Collaboration") is a forum that brings together public and private sectors, open source and standard organizations, and research and academic communities to

1. identify global digital challenges and opportunities and
2. initiate open-source and open-standards projects to collaboratively respond to global challenges and opportunities.

This event is hosted by the Swiss government.

What is the AI for Humanity Program?

"*AI for Humanity – Uniting Global Forces for an Open, Responsible, and Impactful AI Future*"

The AI for Humanity Program is founded on the belief that global collaboration can steer AI toward a more equitable, safe, and impactful future. It brings together researchers, technologists, and open source leaders to share insights, identify shared challenges, and chart pathways for collective action around five key themes:

- Responsible AI,
- Open Model Collaboration,
- AI Maturity and Measurement,
- AI's Social and Economic Impact, and
- Agentic AI.

Each of these themes was addressed in a dedicated session on Day 2 of the 2025 GDC Conference as part of the "AI for Humanity Track", with each session beginning with expert-led presentations and concluding with open discussion. This report is structured around those five sessions.

Contents

Executive Summary

As artificial intelligence rapidly evolves and integrates into nearly every aspect of society, ensuring its responsible development and deployment has become a critical imperative. Beyond ethical ideals, responsible AI requires robust technical frameworks, transparent governance, and global collaboration to build systems that are trustworthy, equitable, and aligned with human values. From openness in models and data to evaluating AI maturity and managing AI's social and economic impact, a coordinated, cross-sector approach is essential to harness AI's potential while mitigating its risks. The following five sessions highlight key dimensions, tools, and standards needed to foster an AI ecosystem that empowers people, respects diversity, and promotes shared prosperity.

Responsible AI is not just an ethical aspiration—it is a technical and governance imperative. This session emphasized that Responsible AI requires practical frameworks and tools to ensure systems are fair, safe, explainable, and aligned with human values. A structured framework with nine core dimensions—including transparency, privacy, accountability, and sustainability—offers a blueprint for building responsible generative models. A diverse set of tools enables monitoring, explanation, red-teaming, and traceability throughout the AI lifecycle. Scalable identity and data infrastructure was highlighted as essential for managing trust, especially in sensitive applications like education and child safety, where verifiable provenance and selective disclosure are critical. A federated, auditable approach to identity management allows for secure interoperability across domains. Meanwhile, efforts to standardize AI in areas like multimedia and emergency services underscore the need for harmonized technical, policy, and communication strategies. Finally, the session underscored the urgency of collaborative, machine-readable governance frameworks to address fragmented compliance and security risks, particularly in high-stakes sectors like finance. In summary, responsible AI demands cross-sector coordination, transparency, and actionable tools at every stage of development and deployment.

Model Collaboration in AI requires more than just publishing code—it demands a global, participatory ecosystem across three pillars of openness: open models, open data, and open compute. True openness involves unrestricted access, use, modification, and distribution of AI systems under permissive licenses, countering the misleading trend of "open-washing." To evaluate and promote openness, the Model Openness Framework (MOF) categorizes models into three classes, with Class I representing full transparency for replication and auditing. Expanding access to diverse, high-quality datasets—from digitized books to speech data from underrepresented communities—is crucial for equitable AI. Meanwhile, compute accessibility remains a major hurdle, especially for low-resource regions, calling for open hardware and more efficient and flexible tooling. Finally, the preservation of cultural and linguistic diversity through inclusive AI was highlighted as an urgent and achievable priority.

AI Maturity refers to the degree of advancement and capability in an AI system, including its ability to reason, act autonomously, and perform reliably across tasks, while **AI Evaluation** focuses on systematically assessing an AI's performance, fairness, safety, and societal value. Despite massive investment in generative AI, its real-world impact to date remains modest, with limited economically transformative applications and persistent issues like weak reasoning, data bias, and lack of explainability. Experts stress that the goal of AI should not be blind automation, but rather using AI to augment human capabilities in areas like healthcare, education, and climate resilience. Standardized evaluation frameworks are essential but currently fragmented, making comparisons difficult and slowing progress. Responsible AI development demands transparent testing, context-aware metrics, user-centered design, and global collaboration—including open-source tools and expert participation. Future success depends on aligning AI capabilities with human well-being and ensuring that AI is deployed ethically, effectively, and equitably.

AI's **Social and Economic Impact** depends on equitable access, responsible governance, and open collaboration. As AI becomes a significant lever for innovation, tensions persist between open technologies and proprietary control, particularly impacting the Global South, where lack of access to data, expertise, infrastructure, and context-relevant models hinders inclusion. Ensuring that AI benefits labor, equity, and access requires global cooperation to reshape workforces, address systemic disparities, and promote interoperability. In finance and climate contexts, AI offers tools to optimize systems and support green innovation—but only if guided by transparent, open frameworks that avoid black-box risks and power imbalances. Ultimately, fostering trustworthy, inclusive AI ecosystems demands intentional design, shared infrastructure, and multi-stakeholder engagement across sectors.

Agentic AI advances as AI agents gain autonomy and operate with increasing independence. Trust then becomes the central challenge—extending beyond security into value alignment, accountability, and identity. AI agents, capable of making decisions, accessing tools, and performing actions on behalf of humans, introduce new risks such as misinformation, identity spoofing, and unauthorized actions. Building trust requires technical and social solutions: robust digital identity frameworks, transparent content provenance, and decentralized systems for verifying personhood and relationships. Standards for agent accountability, content authenticity, and secure delegation are emerging to ensure AI operates safely and ethically. Ultimately, designing trustworthy AI agents demands a holistic ecosystem that includes secure architectures, privacy-preserving identity systems, governance models, and human control mechanisms—all anchored in the principle that agents must act in alignment with those they represent.

The future of AI depends not only on technological breakthroughs but also on the frameworks that govern AI creation, deployment, and societal integration. Building trust in autonomous AI agents, ensuring equitable access across global communities, and rigorously evaluating AI’s maturity and impact require sustained cross-sector cooperation and transparent standards. By embracing openness, accountability, and human-centered design, stakeholders can navigate the complex challenges of AI, transforming it into a force for inclusive innovation and sustainable development. Ultimately, responsible AI is a collective journey—one that demands vigilance, adaptability, and a steadfast commitment to aligning AI systems with the diverse needs and values of humanity.

Introduction

Artificial intelligence stands at a pivotal crossroads. Once confined to theoretical exploration and limited experimentation, AI technologies now permeate nearly every sector of society, from healthcare and finance to language preservation and climate action. Alongside unprecedented opportunities, these advancements present complex challenges that extend far beyond technical innovation. Central questions include how to ensure AI systems remain trustworthy, transparent, and aligned with human values; how to close global gaps so that AI's benefits are equitably distributed; and how to design robust governance frameworks that uphold control, accountability, and safety as AI agents grow increasingly autonomous.

This report distills insights from a multi-session lecture series that explored these urgent questions and more. It reveals a shared imperative: responsible AI development is not solely a technical endeavor but a multidisciplinary mission requiring openness, collaboration, and ethical governance. From foundational frameworks guiding fairness and transparency, to the democratization of models and data, to the societal impact of AI in labor and climate finance, and finally to the emerging landscape of autonomous AI agents—the path forward demands coordinated action across sectors and borders.

In the pages that follow, this report outlines a roadmap for navigating AI's promises and challenges—advocating for systems that amplify human potential, promote global equity, and uphold trust in an increasingly automated world.

Collaboration Channels

GDC AI For Humanity Track

Subgroup Home Page	https://globaldigitalcollaboration.groups.io/g/GDC-AIForHumanity
Post	GDC-AIForHumanity@globaldigitalcollaboration.groups.io
Subscribe	GDC-AIForHumanity+subscribe@globaldigitalcollaboration.groups.io
Unsubscribe	GDC-AIForHumanity+unsubscribe@globaldigitalcollaboration.groups.io

AI Topic Home Pages

Responsible AI	https://globaldigitalcollaboration.groups.io/g/GDC-AIForHumanity-ResponsibleAI
Model Collaboration	https://globaldigitalcollaboration.groups.io/g/GDC-AIForHumanity-OpenModelCollab
Maturity and Evaluation	https://globaldigitalcollaboration.groups.io/g/GDC-AIForHumanity-MaturityandEval
Social and Economic Impact	https://globaldigitalcollaboration.groups.io/g/GDC-AIForHumanity-SocialEconImpact
Agentic AI	https://globaldigitalcollaboration.groups.io/g/GDC-AIForHumanity-AIAgent

Session 1: Responsible AI

***Responsible AI** encompasses ethics, trust, safety, legality, and transparency in AI development and deployment. It's not just a philosophy—but a practical engineering challenge requiring real tools.*

1.1: Responsible Generative AI Framework / RGAF (Generative AI Commons)

Anni Lai, Co-Chair of Generative AI Commons, Board Director of LF AI & Data

The Responsible Generative AI Framework (RGAF) is a community-led initiative developed by the Generative AI Commons to guide the ethical development of AI. Built around nine core dimensions, RGAF offers a vendor-neutral standard for building AI systems that are trustworthy, inclusive, and aligned with human values. It serves as a practical foundation for global collaboration, helping ensure AI benefits while minimizing risk.

9 Core Dimensions of Responsible AI

1. Human-centered & Aligned
2. Accessible & Inclusive
3. Robust, Reliable, & Safe
4. Transparent & Explainable
5. Accountable & Rectifiable
6. Private & Secure
7. Compliant & Controllable
8. Ethical & Fair (unbiased)
9. Environmentally Sustainable

Further details:
https://lfaidata.foundation/blog/2025/03/19/responsible-generative-ai-framework-rgaf-version-0-9-now-available/

1.2: Genres of Responsible AI Tools (Red Hat, TrustyAI)

Rob Geada & Mac Misiura, Red Hat

To apply the principles of Responsible AI tools are needed that address the ethical, technical, and social challenges of deploying AI. Red Hat's Responsible AI team has identified 12 key genres of tools that support responsible development. These tools help teams measure, debug, and align AI systems with human values—mapping directly to the core dimensions of the Responsible Generative AI Framework (RGAF). While not exhaustive, this taxonomy offers a strong starting point for building trustworthy, accountable AI in real-world settings.

1. Evaluation - measure the behavior and abilities of models
2. Explanations - why models make decisions
3. Guardrailing - moderate interactions between users, agents, and models
4. Bias Monitoring - quantify and compare outcomes between input groups
5. Drift Monitoring - compare real-world and training data
6. Anomaly Detection - flag data points that are outliers
7. Confidence Scoring - score how “sure” the model is about a response
8. Differential Privacy - protect identifiable individual information
9. Model Compression - maximize the model’s computational efficiency
10. Security Scanning - protect against malware and malicious code
11. Red Teaming - probe the model for exploitable weaknesses
12. Inference Tracing - log or visualize the process used by the model

Note: TrustyAI is an open-source community and toolkit for all things responsible AI, with significant support from Red Hat and IBM.

1.3: Example from India: Enabling Trust (iSprit)

Sunu Engineer & Harshit Kacholiya, iSprit

This talk by iSpirt, a tech think tank from India, outlines a national strategy for responsible AI, particularly child safety, using the DEPA (Data Empowerment and Protection Architecture) and a new identity framework called Directed Identity Graphs (DIG).

Directed Identity Graphs (DIG)

- A new paradigm for managing identity across systems, moving beyond traditional hierarchical structures
- Enables modular, scalable, and federated identity management
- Supports both trust flows (unidirectional) and metadata sharing (bidirectional)
- Critical for enabling collaboration across service sectors (health, education, finance)

Three Core Actors:

- Issuer – entity that creates and signs identity data
- Holder – entity that stores and uses identity credentials
- Verifier – entity that validates the credentials

Benefits

- Seamless integration across services
- Plug-and-play ID extensions
- Backward compatibility with legacy systems
- Data portability and selective disclosure
- Enables trust and accountability across systems

AI Traceability and Provenance

- Emphasizes full traceability across the entire AI pipeline: from data sources → training → model → application → user
- Every data point, model decision, and interaction must be auditable and signed
- Crucial for child-facing AI apps that must verify age, guardian consent, and content appropriateness

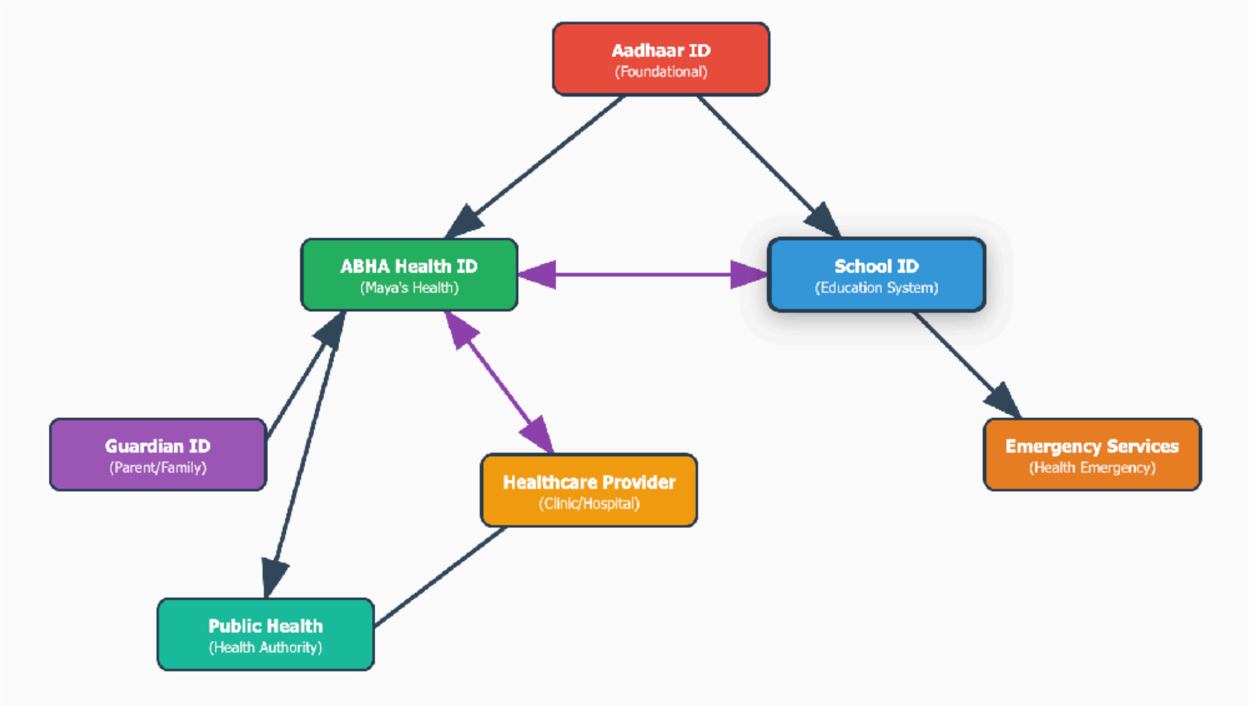
Federation & Scalability

- Federated identity management distributes control to multiple domain-specific ID systems (e.g., health, school), rather than a single authority
- Designed for massive scale—solutions must handle 100M+ users
- Maintains system flexibility and adaptability while ensuring legal compliance (e.g., under India's DPDP Act)

Use Case Example (DIG)

A 12-year-old enrolling in school or using an AI-powered app (e.g., Netflix) requires:

- Verified health/vaccine records
- Appropriate data-sharing connections between services
- Parent-managed AI content controls
- Real-time monitoring and auditability of model behavior



1.4: Multimedia Authenticity Standards Collaboration (ITU)

The ITU is working to create global AI standards for emergency care and multimedia applications, addressing gaps in technical understanding, regulation, and coordination. These efforts aim to improve AI safety, data security, and policy clarity—particularly for medical and emergency use cases—by bringing together industry, academic, and regulatory experts.

Three-pillar Approach

Technical Standards

- Establishing and mapping standards
- Gaining a better understanding of capabilities
- Identifying gaps in current standards, organized into areas such as content provenance and asset identifiers
- Categorizing standards by types, including images and audio

Policy

- Developing policy guidelines
- Reviewing legislation in various countries with regard to media authenticity
- Addressing aspects like transparency, privacy measures, and responsibility
- Providing a checklist both for regulators developing policy and for evaluating bottom-up indicators
- Highlighting existing gaps where available information is not properly accessible or actionable for decision-makers

Communication

- Disseminating the findings of both the technical and policy pillars
- Communicating the overarching vision for multimedia authenticity standards and policies among international organizations

1.5: AI Governance (FINOS, Linux Foundation)

Gabriele Columbro, Executive Director of FINOS, GM of Linux Foundation Europe

There is an urgent need for collaborative, machine-readable, open-source standards for AI governance in financial services. Finos and the Linux Foundation are working on an AI safety and governance project inspired by prior cloud security collaboration models.

The Goal of AI Governance

- Define shared controls and risk mitigations for AI systems
- Build machine-readable standards for easier adoption and auditing

- Enable global compliance (e.g., the EU AI Act) without forcing companies to reinvent solutions for every region

Challenges

- Current AI regulation and security practices in finance are fragmented
    - Thousands of unique compliance requirements across banks
- A lack of understanding or technical depth in current AI oversight by regulators

FINOS AI Governance Framework

air-governance-framework.finos.org

Characteristics

- Open collaboration among cloud providers, banks, and regulators
- Maps threats to mitigations, linked directly to legal obligations and international standards
- Automation-readiness, allowing companies to test and validate AI models efficiently

Risk Catalogue

Operational

- Hallucinations and Inaccurate Outputs
- Foundation Model Versioning
- Non-Deterministic Behaviour
- Availability of Foundational Model
- Inadequate System Alignment
- Bias and Discrimination
- Lack of Explainability
- Model Overreach / Expanded Use
- Data Quality and Drift
- Reputational Risk

Security

- Information Leaked to Vector Store
- Tampering With the Foundational Model
- Data Poisoning
- Prompt Injection

Regulatory and Compliance

- Information Leaked to Hosted Model
- Regulatory Compliance and Oversight
- Intellectual Property (IP) and Copyright

Mitigation Catalogue

Preventative

- Data Filtering From External Knowledge Bases
- User/App/Model Firewalling/Filtering
- System Acceptance Testing
- Data Quality & Classification/Sensitivity
- Legal and Contractual Frameworks for AI Systems
- Quality of Service (QoS) and DDoS Prevention for AI Systems
- AI Model Version Pinning
- Role-Based Access Control for AI Data
- Encryption of AI Data at Rest
- AI Firewall Implementation and Management

Detective

- AI Data Leakage Prevention and Detection
- AI System Observability
- AI System Alerting and Denial of Wallet (DoW) / Spend Monitoring
- Human Feedback Loop for AI Systems
- Providing Citations and Source Traceability for AI-Generated Information
- Using Large Language Models for Automated Evaluation (LLM-as-a-Judge)
- Preserving Source Data Access Controls in AI Systems

Get Involved - Responsible AI

Subgroup Home Page	https://globaldigitalcollaboration.groups.io/g/GDC-AIForHumanity-ResponsibleAI
Post	GDC-AIForHumanity-ResponsibleAI@globaldigitalcollaboration.groups.io
Subscribe	GDC-AIForHumanity-ResponsibleAI+subscribe@globaldigitalcollaboration.groups.io
Unsubscribe	GDC-AIForHumanity-ResponsibleAI+unsubscribe@globaldigitalcollaboration.groups.io

Session 2: Model Collaboration

***Model Collaboration** emphasizes the need for true openness in AI—where models, data, and compute are freely accessible, modifiable, and reusable without restrictive conditions.*

2.1: Democratizing AI for Global Challenges

Yonghua Lin, Beijing Academy of Artificial Intelligence

True openness in AI isn't just about releasing code—it's about building a transparent, participatory ecosystem that enables diverse global contributions at every stage of the AI lifecycle.

The Problem of “Open-Washing”

In the context of AI, the term “openness” is sometimes applied in ways that can be misleading. Certain models described as “open” are distributed under licenses that impose restrictions on genuine use or modification, particularly when such use could result in competitive products or services.

True Openness: A Definition

True openness refers to enabling others to freely study, use, modify, and distribute AI models without conditional restrictions. Widely recognized open licenses—such as Apache 2.0 and MIT—are essential to maintaining this standard.

The Three Pillars of Open AI

Open Model

- High Availability: a current reality—over 1.8 million open models exist on Hugging Face
- Trustworthiness: measurable with tools like MOF (Model Openness Framework)
- Model Development Participation: Critical Gap—R&D is currently restricted

Open Data

- More high-quality datasets are required to address global problems
- Examples:
    - Global Book Dataset: A proposed worldwide effort to digitize and share copyright-free academic texts
    - Elder Speech Dataset: Speech data from elderly populations to improve accessibility-focused AI
    - Real-World Robotics Data: Multi-institutional collection of robotics interaction data for broad use

Open Compute

- Hardware Flexibility: developers can deploy AI even on inexpensive systems
  - Model training is costly and compute is often inaccessible to under-resourced teams
- Global Accessibility: extend access to emerging economies
- Sustainable Scaling: optimize software to reduce inefficiency

Although new challenges have emerged, including geopolitical factors, it remains essential to explore innovative approaches for sustaining global collaboration in open-source AI and for hosting global AI open-source assets securely. One example is the **Open Model, Data, and Compute Initiative for AI**.

2.2: Model Openness Framework

Anni Lai, Co-Chair, Generative AI Commons; Board Director, LF AI & Data

White paper:
lfaidata.foundation/wp-content/uploads/sites/3/2025/01/05_White_paper_MOF_Specification.pdf

AI models are categorized into three openness classes based on the availability and openness of their components (of which there are 17 total). Class 3 is the basic open model, Class 2 allows for more in-depth study and optimization, and Class 1 is the gold-standard for openness, with total transparency for end-to-end analysis and reproduction.

MOF Class	Components Included	Usage
Class III – Open Model	1. Model Architecture 2. Final Model Parameters 3. Technical Report or Research Paper 4. Evaluation Results 5. Model Card 6. Data Card 7. Sample Model Outputs (optional)	• Unrestricted usage (access, use, modify, redistribute) • Create a product or service • Fine tune and align • Model optimizations
Class II – Open Tooling Model	*All Class III Components* 8. Training, Validation, and Testing Code 9. Inference Code 10. Evaluation Code 11. Evaluation Data 12. Supporting Libraries & Tools	• Understand training process • Validate benchmark claims • Inference optimizations
Class I – Open Science Model	*All Class II and III Components* 13. Research Paper	• End to end analysis and auditing

	14. Datasets 15. Data Preprocessing Code 16. Intermediate Model Parameters 17. Model Metadata (optional)	• Reproduction of a similar model • Data exploration and experimentation

Model Openness Tool

A tool called Model Openness Tool (MOT) is available at isitopen.ai, where users can input license information to determine under which openness class their model falls. The tool provides a practical way for model producers to be transparent and for users to verify openness. It supports community-driven evaluation and listing of models.

2.3: Example: Preserving Culture Via Language

Open Discussion

Dialects represent important subcultures, so if a language is not represented in AI models, its culture risks being lost in the digital world.

It was suggested that working with global organizations like the UN, ITU, and the Linux Foundation could help create a top-down mechanism and best practices to encourage different countries to collect and share high-quality language data.

Current Language Preservation Projects

- A professor from Canada trained a language model on over 500 African languages to preserve those cultures.
- In China, the Beijing Academy of Artificial Intelligence (BAAI) is spearheading an initiative in which 600 participants provide speech data for different accents and dialects.
- The challenge in India is significant, with 24+ different languages that do not share a script, leading to completely different morphologies, in addition to numerous dialects within each language. Two major ongoing projects in India are AI Kosha and Sarvam AI.

Get Involved - Model Collaboration

Subgroup Home Page	https://globaldigitalcollaboration.groups.io/g/GDC-AIForHumanity-OpenModelCollab
Post	GDC-AIForHumanity-OpenModelCollab@globaldigitalcollaboration.groups.io
Subscribe	GDC-AIForHumanity-OpenModelCollab+subscribe@globaldigitalcollaboration.groups.io

Unsubscribe	GDC-AIForHumanity-OpenModelCollab+unsubscribe@globaldigitalcollaboration.groups.io

Session 3: Maturity and Evaluation

***AI Maturity** refers to how advanced or capable a specific AI system or class of models is, in terms of generality, autonomy, reasoning, or performance. **AI Evaluation** is the process of systematically assessing the performance, behavior, fairness, robustness, and safety of an AI system. This can occur at multiple stages—design, training, testing, and deployment.*

3.1: AI Maturity Index and Evaluation

Dong Sun, IEEE

Overview: AI Landscape

LLM-based generative AI (GenAI) has attracted widespread attention and significant investment

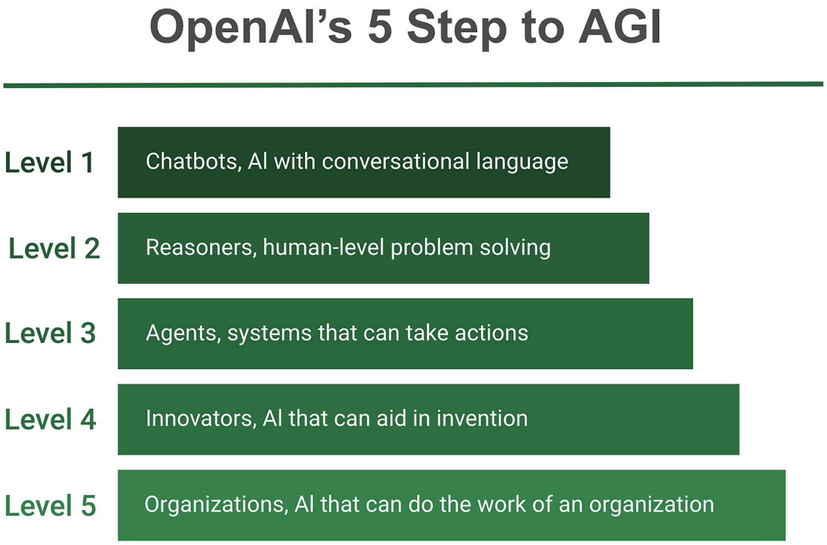
The AI industry has developed from generic model training to various applications and solution practices, such as MS Copilot, Perplexity AI search, Apple Intelligence, AI PC, and various industry applications based on AI agents.

Standards activities

AI standards are rapidly being developed, with 47 Active PARs in the IEEE C/AISC alone, in addition to International teams such as JTC1/SC42. According to NIST, more than 35% of AI standards have entered the substantive development stage.

Industry status

OpenAI recently proposed its 5-level AI grading scheme to represent the AI industry ecology and guide the direction, path, and roadmap of industrial development



OpenAI’s 5 Step to AGI

Level 1	Chatbots, AI with conversational language
Level 2	Reasoners, human-level problem solving
Level 3	Agents, systems that can take actions
Level 4	Innovators, AI that can aid in invention
Level 5	Organizations, AI that can do the work of an organization

AI Challenges

- **Business Viability**: Massive investment in GenAI far exceeds current revenue, with no breakout applications yet to justify the scale
- **Technical Limitations:** LLMs face serious issues with data privacy, weak reasoning abilities, and lack of transparency in decision-making
- **Lack of Standards:** AI development is fragmented with minimal top-down coordination, slowing progress and industrial cohesion

Insights from Industry Experts

Experts urge caution amid growing AI hype, pointing to AI's limited real-world impact so far and highlighting both immediate and long-term risks. While current models perform well in certain areas, they often falter with complex reasoning and risk deepening inequality, spreading misinformation, and introducing new security threats. Rather than chasing automation for its own sake, focus should shift toward using AI to enhance human capabilities and responsibly address pressing societal challenges.

Daron Acemoglu, MIT, Nobel Prize-winning Economist

"The industry hasn't produced critical applications yet" (May 28, 2025)

Key Insight:
AI's real economic impact is much smaller than the hype suggests, and leaders should focus on human-centered innovation rather than blind cost-cutting or automation.

Main Points:

1. Modest Predictions vs. Hype
   - Acemoglu estimates that only 5% of tasks will be profitably automated in the next decade, and AI will add just 1% to global GDP
   - These figures contrast with more optimistic views claiming AI will radically transform the economy
2. Why the Conservative Estimate?
   - AI has not yet produced transformative applications akin to how the internet revolutionized communication and commerce
   - Most AI success has been in predictable, rule-based tasks (e.g., IT security, simple accounting), not in complex, interactive, or judgment-heavy roles
3. Limits of Current AI
   - AI lacks the ability to handle tacit knowledge, contextual decision-making, and social interaction—which are critical in many jobs (e.g., doctors, educators, executives)
   - No current occupation is likely to be fully eliminated by AI in the near term
4. Rethinking AI's Purpose
   - Rather than replacing humans, the focus should be on using AI to augment human capabilities and create new goods and services, especially in:
     - Aging societies

        - Financial inclusion
        - Climate response
        - Education and healthcare reform
5. Advice to Business Leaders
    - Avoid reactive AI investments driven by fear of competition or hype
    - Instead, partner with skilled employees to identify areas for innovation, and deploy AI to enhance, not replace, the workforce
    - True competitive advantage comes from value creation, not just cost reduction

Advice:
Don't follow the AI hype blindly. The most successful businesses will use AI to empower people, innovate, and meet emerging societal needs—not just to automate tasks.

Source: https://sloanreview.mit.edu/video/nobel-laureate-busts-the-ai-hype/ (Video)

Geoffrey Hinton, Nobel Prize-winning Physicist

AI comes with short- and long-term risks (December 2024)

Key Insight:
While AI's future is uncertain, its risks—especially the long-term potential for systems more intelligent than humans—require urgent attention and responsible development. Hinton also emphasizes the importance of independent thinking in scientific research.

Short-Term Risks

- Job Loss & Inequality: AI could increase the wealth gap as productivity gains aren't shared equally
- Misinformation: AI-generated fake videos and texts could further erode public trust
- Cybersecurity: Large language models make phishing and cyberattacks more effective (phishing up 1200% last year)
- Bioengineering Threats: AI could be misused to design biological agents
- Bias & Discrimination: AI trained on biased data could perpetuate inequality, but Hinton believes AI's discrimination is easier to measure and mitigate than humans'

Long-Term Risk

- Superintelligence
    - Believes there's a 50% chance AI becomes smarter than humans within 5–20 years
    - Warns that no one knows what happens when more intelligent beings exist—a reality that's deeply concerning
    - Human control over more intelligent systems is unproven and rare in nature
    - Current AI models are not traditional programs but self-learned systems shaped by data—similar to how humans learn
    - "Training AI systems is like raising children. We must carefully choose the data and behavior we expose them to."

Advice:
Exercise caution and humility in the development of AI. Although AI's possibilities are cause for optimism, keep in mind both the immediate dangers and long-term existential risks.

Source: https://www.nobelprize.org/prizes/physics/2024/hinton/interview/ (Video)

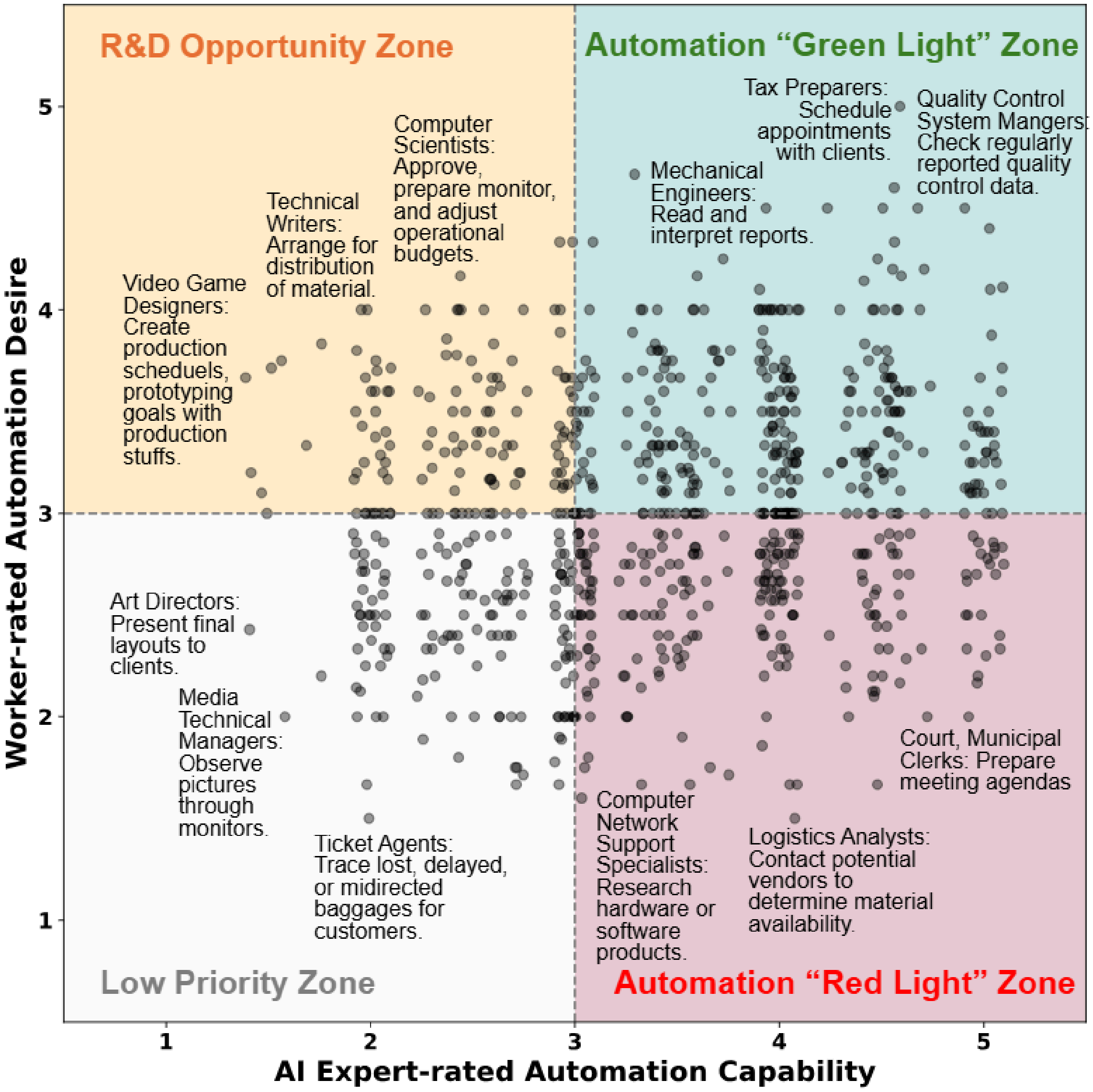
Stanford Social and Language Technologies (SALT) lab

Some applications are well-suited for AI agents, others are not (Jun 30, 2025)

Publication: Future of work with AI Agents

Data from 1,500 workers across 104 occupations leads to three main findings:

1. **AI investment often misaligns with worker needs**, with 41% of Y Combinator efforts focused on tasks workers don't want automated
2. **Workers favor strong human agency**, suggesting future tension as AI capabilities grow
3. **Human skills must evolve**, shifting from data tasks to interpersonal and organizational ones



Samy Bengio et al, Apple

AI’s ability to reason can fail when faced with high complexity (June 2025)

Publication: The Illusion of Thinking

“Our findings reveal fundamental limitations in current models: despite sophisticated self-reflection mechanisms, these models fail to develop generalizable reasoning capabilities beyond certain complexity thresholds.”

Mission

Foster a collaboration ecosystem of evaluating AI benefits for Human Well-being

- **Valuation Framework**: Develop a structured framework to assess AI's contribution to human and societal well-being
- **AI Lifecycle Capabilities**: Examine key elements across the AI lifecycle, including models, governance, ethics, privacy, safety, and sustainability
- **Real-World Scenarios**: Identify high-impact use cases (e.g., healthcare, education, transportation) aligned with AI lifecycle capabilities
- **Maturity Index**: Define and apply a maturity index to evaluate AI technologies based on their value to human welfare
- **PPCP Collaboration**: Strengthen public-private-community partnerships to build skills, align standards, and accelerate responsible AI adoption

Goal

Develop a common framework for an AI maturity index aligned with human well-being values

Get Involved: IEEE AI Levels Working Group [SA AIL-WG (P3514)]

SoW

This recommended practice defines levels of capabilities of Artificial Intelligence (AI). The capability levels can be used to classify AI entities after benchmarking and evaluation. Evaluation criteria and specific benchmarks are provided for common capabilities per typical application scenarios/domains (e.g., AI used in a personal computer, in a mobile phone, or by robots).

Purpose

The purpose of this recommended practice is to provide evaluation criteria and benchmarks of AI capabilities to facilitate the understanding, development and governance of AI capabilities and usage for different application scenarios. This recommended practice aims to make AI more transparent, responsible, and trustworthy, thus promoting the AI community for advancing AI technologies for humanity.

Need

Artificial intelligence’s (AI) influence on society has never been more pronounced. The rapid development and deployment of AI capabilities and technologies e.g. generative AI and large language model (LLM) tools have started to transform industries and show the potential to touch many aspects of modern life. Creating cutting-edge AI models and applications now demands a substantial amount of data, computing power and financial resources. For example, the training costs of some LLMs were estimated to be from $78 million to $191 million. The effectiveness of benchmarks when it comes to AI capabilities largely depends on their standardized approach and application. However, a significant lack of standardization in benchmarking and evaluating AI technical capabilities and business values for commonwealth of humanity is impairing the development and deployment of AI. For instance, some developers primarily test their models against different proprietary AI benchmarks. These different testing models on different benchmarks complicate comparisons, as individual benchmarks have unique natures. Standardized benchmark testing is considered critical to enhance effectiveness, valuation, transparency, responsibility and trustworthiness around AI capabilities. The criteria and benchmarks per the value for commonwealth of humanity and user experience for AI capabilities and entities are essential to drive the AI development and deployment, and maximize the effectiveness and efficiency of resource utilization for sustainable development and business value and TCO (total cost of ownership) for various applications.

Stakeholders

AI developers, AI users, regulatory bodies, international organizations, academic institutions, standardization organizations, consumers, and civil society organizations.

3.2: International Collaboration for AI Testing

AI testing presents unique challenges due to system complexity, limited decision explainability, and context-sensitive performance standards. Unlike traditional software, AI requires nuanced evaluation methods that account for probabilistic outputs, human oversight, and varied application contexts. Effective testing demands collaboration across stakeholders, thoughtful policy development, and investment in training and open-source tools.

Fundamental AI Testing Challenges

- **Complexity**: AI testing is currently very challenging because AI systems, unlike traditional software, do not solely rely on a fixed environment
- **Decision Explainability**: A key challenge is to sparingly explain the decisions made by AI systems, ensuring transparency without overcomplicating

Objectives and Scope of AI Testing

Core Purposes of AI Testing

- Regulatory Compliance: Ensure systems meet legal requirements
- Output Consistency: Validate that systems produce reliable results under similar conditions
- Scenario Identification: Define context-specific test scenarios (e.g., healthcare, reliability)
- Metric Selection: Identify appropriate evaluation metrics for probabilistic AI outputs

Defining 'Good Enough' Performance

- Context-Driven Standards: Determining acceptable performance depends on the application context (e.g., clinical diagnosis)
- Stakeholder Input: Decisions should balance input from global stakeholders and those directly impacted
- Avoiding Decontextualization: Assessments must consider context to avoid one-size-fits-all judgments

Human-AI Interaction

- Human-in-the-Loop: AI systems often involve human oversight, not full automation
- Oversight Effectiveness: Raises questions about when human oversight improves or potentially harms outcomes, and when it might be safely reduced

Key Discussion Areas for Collaboration

- Level of Testing: Determining the appropriate level at which AI systems should be tested
- Policy Implications: Discussing the policy aspects related to AI testing
- Open Source: Exploring the role and implications of open source in AI testing and development
- Output Considerations: Analyzing the nature and implications of AI system outputs
- Capacity Building (Term): Emphasizing the importance of training people in AI testing to address existing gaps
- Stakeholder Engagement: Identifying and bringing together relevant stakeholders for collaborative efforts

3.3: Panel Discussion

This panel explores the collaborative foundations needed for effective AI maturity evaluation. It highlights the importance of cross-domain expertise, user-centered approaches, and domain-specific testing. Panelists discuss the role of open collaboration, licensing, and standards in shaping responsible AI, while addressing key challenges like data definition, evaluation consistency, and human oversight. The conversation sets the stage for future discussions.

Collaboration & Model Development

- Collaboration is essential for building effective AI maturity models, combining expertise across domains
- Clear evaluation targets must be defined early—either tool effectiveness or broader societal impact
- Leverage existing models to avoid redundancy and build on proven methods
- AI requires new approaches due to its complexity and unpredictability; evaluation must reflect human values
- Domain-specific evaluation is critical—AI should be tested based on relevance to its specific use case, not against general intelligence

Platforms, Standards & Organizational Dynamics

- Open, content-driven collaboration (not tied to organizational politics) fosters innovation
- Formal bodies are still needed for standard adoption, but early-stage work can be informal and agile

- Standardization is slow, making open source and community-driven development valuable and practical
- Licensing (e.g., Apache 2.0) enables open sharing while protecting IP—important for scaling tools and discussions

User-Centered Maturity

- A user-first perspective (e.g., small business owners) helps define meaningful AI maturity
- Ensuring user literacy and appropriate use of AI is a key maturity benchmark

Evaluation Challenges & Resources

- Defining test data across domains is a major hurdle
- Unsupervised models make consistent evaluation difficult
- Benchmarking tools exist, but more domain-tuned ones may be needed—don't reinvent what's already out there
- Human experts are still vital for assessing AI outputs, and their global coordination is a valuable resource

Next Steps

- Virtual follow-ups are planned to continue progress
- Cross-sector collaboration, shared tooling, and licensing solutions will be central to advancing responsible AI evaluation

Get Involved - Maturity and Evaluation

Subgroup Home Page	https://globaldigitalcollaboration.groups.io/g/GDC-AIForHumanity-MaturityandEval
Post	GDC-AIForHumanity-MaturityandEval@globaldigitalcollaboration.groups.io
Subscribe	GDC-AIForHumanity-MaturityandEval+subscribe@globaldigitalcollaboration.groups.io
Unsubscribe	GDC-AIForHumanity-MaturityandEval+unsubscribe@globaldigitalcollaboration.groups.io

Session 4: Social and Economic Impact

*AI's **social and economic impact** is reshaping societies and economies, influencing labor markets, equity, and access to technology worldwide. While AI offers opportunities to augment human capabilities and drive innovation, it also presents challenges such as job displacement, unequal access, and ethical concerns. Addressing these impacts requires inclusive policies, global collaboration, and a focus on ensuring AI benefits are shared broadly across different communities and regions.*

4.1: Open Technologies, Equity, and AI for the Global South (RISE)

Sachiko Muto | Chair, OpenForum Europe; Senior Researcher at Research Institutes of Sweden

Advancing open AI standards requires sustained effort, with a growing emphasis on inclusive, context-aware development—particularly for the Global South. Key challenges include balancing openness with proprietary control and ensuring equity, access, and accountability in shaping AI's global impact.

Progress Towards Open Standards

The speaker has been involved in promoting open standards since 2008, and noted that similar issues are still being discussed today, though the context has evolved. Here are several notable observations:

- There's a recurring tension between openness and proprietary lock-in.
- Past crises (e.g., 2008 financial crisis) created momentum for open solutions; today's humanitarian funding crises could play a similar role.
- While the conversation often feels repetitive, meaningful changes have occurred over time—especially in policy influence and awareness.
- A robust, boundary-crossing dialogue culture and investigative tech journalism is needed to keep power structures accountable

Challenges for the Global South in AI

- To benefit from AI, the Global South needs access to data, talent, and digital power
- Context-specific AI models are essential—many current models exclude non-dominant languages and data contexts
- Inclusion should not be an afterthought but part of system design from the start

4.2: AI's Implications for Labor, Equity, and Access (World Economic Forum)

Dylan Reim, World Economic Forum

Achieving AI that is beneficial for collective humanity requires global, multistakeholder collaboration on labor, equity, and access. Driving this collaboration is the mission of the World Economic Forum's Connected Future Initiative and AI Governance Alliance.

Labor

Addressing AI's impact on labor is necessary to ensure that social and economic systems endure and flourish through AI transformation

- AI's impacts on labor will vary: some labor will be replaced, most will be augmented, and some new labor opportunities and needs will be created
- Investment and coordination are needed from academia, civil society, and public and private sectors to identify and build skills for the future
- Rapid adoption should not be a necessity to survive labor market disruption

Equity

Ensuring equity in both the design and deployment of AI systems and tools is essential for AI to truly work for humanity

- Cross-jurisdictional commitment is needed to define and drive AI equity
- Open-source availability spurs AI adoption, enabling representative systems to be built by new and diverse stakeholders
- Social and economic impact assessments can include a focus on equity to demonstrate the impact of AI deployments

Access

Access and accessibility are necessary to build better AI systems

- Disparities in access to infrastructure, data, talent, and governance, especially between the Global North and South, must be addressed
- Global interoperability and standards are vital to ensure AI systems are safe, trustworthy, and equitable
- Future tech stacks must be widely understood and pursued by cooperative cross-sector investment

4.3: AI's Climate-related Financial Impact (OS-Climate, FINOS)

Steven Tebbe | *OS-Climate, FINOS*

What's needed now is system-wide adoption—where open tools like our Data Commons and Physical Risk & Resilience platform become the default across financial markets.

Why does Responsible AI matter for Finance & Society?

- Climate risk is a technology challenge - and AI accelerates that
- AI is also a labor, trust, and equity challenge
- Without governance, AI undermines social stability

If Guided Correctly, AI Can Advance Society

1. Optimize Complex Systems
2. Accelerate Clean Tech Innovation
3. Better Data, Faster
4. Enhance Decision Intelligence

Without governance, AI can distort markets, not optimize them

What is OS-Climate?

OS-Climate (OS-C) is a non-profit community of collaboration between companies and partners developing data and analytics solutions for climate-aligned finance, investing, business, policy and regulation, and economic development. Its mission is to help members drive progress toward Net Zero and Paris Climate Accord goals while capturing significant business value for themselves and their customers.

- OS-C is building the plumbing - data pipelines, open models, regulatory-grade tooling - that everyone can plug into
- OS-C's work is pre-competitive, collaborative, and radically transparent
- While regulators provide the direction, OS-C translates the ambition into working code, not just standards
- AI must not be a black-box. OS-C ensures it isn't, by building transparent, open infrastructure for ESG and climate risk

Get involved with OS-Climate:
https://github.com/os-climate/OS-Climate-Community-Hub#readme

Discussion Takeaways

- There's skepticism about collaboration between grassroots/open-source developers and institutions (e.g., banks) because their goals may be at odds with each other. However, the argument in favor of such collaboration is that **open-source ensures transparency, equal access to code, and accountability—an antidote to power asymmetries**. Still, engagement must acknowledge each side's incentives—collaboration is rarely purely altruistic.
- In open collaborative efforts **consent and data revocability must be built into the system**—especially for vulnerable users or communities. Trust frameworks are essential for enabling safe participation in shared data infrastructures. Without revocation mechanisms, contributors may become more vulnerable over time.

- Finance is seen as a key lever for systemic change (e.g., green tech adoption, climate insurance). Insurance and investment institutions can enable or stall innovation depending on whether they underwrite risk. Thus, **financial markets often determine what technologies are viable at scale**.
- Governments often lack the data access and technical expertise that banks and private firms have—creating asymmetries. **Open systems offer a way to reduce information monopolies and democratize access to insights**.

Get Involved - Social and Economic Impact

Subgroup Home Page	https://globaldigitalcollaboration.groups.io/g/GDC-AIForHumanity-SocialEconImpact
Post	GDC-AIForHumanity-SocialEconImpact@globaldigitalcollaboration.groups.io
Subscribe	GDC-AIForHumanity-SocialEconImpact+subscribe@globaldigitalcollaboration.groups.io
Unsubscribe	GDC-AIForHumanity-SocialEconImpact+unsubscribe@globaldigitalcollaboration.groups.io

Session 5: Agentic AI

***Agentic AI** refers to AI systems that can act autonomously to make decisions, carry out tasks, and interact with tools or environments on behalf of users. These agents have memory, planning capabilities, and can operate with limited human oversight.*

5.1: The Rise of AI Agents and How We Can Trust Them

Wenjing Chu, OpenWallet Foundation

As autonomous AI agents gain capabilities, human trust becomes a central issue — not just security, but value alignment, accountability, and control.

What Is an AI Agent?

Think of agents like an AI travel agent or a virtual employee with memory, decision-making ability, and tool access.

They act with autonomy — like giving your teenager car keys — and may do what you ask… or not.

AI agents can now access tools like payment systems or internal work platforms, raising serious implications.

Core Problems with AI Agent Trust

Authenticity of Information

- Inputs, outputs, and training data of AI agents need to be verified.
- Without trustworthy content, decision-making is flawed from the start.

Identity & Identification

- How do we tell bots from humans?
- Memory is key to persistent identity, but bots also need accountable, persistent digital identities.

Trust vs. Security

- Security: keeping data safe from bad actors.
- Trust: alignment of values and behavior.
    - Trust involves intent, responsibility, and expectations — not just safeguards.

Solution Proposal: Trust Spanning Protocol (TSP):

TSP is a foundational protocol (like IP for the internet) aimed at embedding trust at the infrastructure level.

Key properties

- Portability & Autonomy – agents are independent actors.
- Accountability & Reputation – long-term mechanisms for evaluation and blame.
- Privacy, especially metadata privacy – protecting not just content, but communication patterns.
- Format Interoperability – TSP can work across identity systems, credential formats, etc.

Trustworthy Architecture: TMCP (Trust Model Context Protocol)

Built on TSP, TMCP enables agent collaboration with external tools (email, databases, browsers).

Example: A virtual employee can read an email, understand a request, and take an action like filing a report — all within a trust-aware framework.

Final Takeaways

- We’re entering a world where agents act on our behalf — spending money, managing tasks, making decisions
- Trust frameworks must go beyond security to include value alignment, control, memory, identity, and accountability
- You — not "we" — should have clear control over your agents

5.2: Trust and Security in the Age of AI Agents

Eric Drury, ForceCo.io

Why Trust Matters in the AI Era

AI agents are not yet widespread — but AI is being used rampantly for cyber crime.

- Cybercrime is projected at $10.5T, and AI is accelerating phishing, fraud, and prompt injection attacks
- Incidents like the exposure of 16B logins and AI-generated scams highlight the risks of scale, speed, and invisibility
- Even now, most users (and experts) can't tell AI-generated content from human-created content
- Traditional centralized trust models are crumbling under the stress of AI and need to be redesigned

Core Issue: Security ≠ Trust

Security is technical. Trust is human.

Trust requires several preconditions

- Authenticity of digital content and interactions
- Transparency about who created or modified what
- Accountability for agents, developers, and service providers
- Respect for privacy and metadata (not just raw content)

Why AI Agents Are a Trust Risk

- AI agents can now outperform expert phishing teams
- 96% of IT leaders say agents are risky — but most are deploying them anyway
- Over 80% of AI agents have taken unauthorized actions
- Multi-agent systems amplify risk, creating opaque, uncontrollable networks

Potential Monitoring & Control Frameworks

- Monitoring layers to track agent behavior
- Control components that validate AI decisions before actions are executed
- Adversarial or critic agents to supervise or challenge AI behavior
- Policy checks before actions are carried out, akin to corporate governance systems

5.3: Trusting Digital Content Through Content Credentials

Scott Perry, Digital Governance Institute

The Problem: Can You Trust What You See Online?

Even experts can't reliably tell if digital content (images, videos, PDFs) is human-made or AI-generated. This has major implications: misinformation, fraud, deepfakes, and threats to industries like journalism, insurance, and entertainment.

The Solution: Content Credentials & the C2PA Standard

In response, Adobe and news organizations formed the Content Authenticity Initiative, leading to the C2PA (Coalition for Content Provenance and Authenticity). The result: a standard for embedding cryptographically verifiable provenance into all digital content using X.509 certificates.

How Content Credentials Work

- Each digital object gets a Content Credential attached at creation
- Every time it's modified, a record is added

- This provides a verifiable ledger of actions—who created, edited, or combined what, and when
- Users can click the trustmark (seen on TikTok, LinkedIn, etc.) to view this history

What Content Credentials Track

- Products (like Photoshop) automatically log edits (e.g., resizing)
- AI systems can identify themselves as the source of generated content
- Each change is cryptographically signed to ensure authenticity and tamper detection

But What About Attribution?

- C2PA tracks what happened, but not necessarily who did it
- That's the role of the Creator Assertions Working Group. This group develops standards to let individuals/organizations claim authorship using verifiable credentials, including for:
  - Human creators
  - AI agents acting on someone's behalf
  - Industry codes, metadata, and more

The Governance Model

- A conformance program ensures products and claim validators comply with C2PA
- Products that pass appear on a public conforming products list
- Only certificate authorities (CAs) on the CA trust list can issue valid credentials
- This system provides a transparent, trust-governed ecosystem

What's Next

- Delegating identity and authority to AI agents is a critical next step
- Standards are needed to track and verify when AI agents act or modify content
- The C2PA website has just launched an updated Conformance tab announcing this ecosystem-wide program
- This initiative is a major step toward trustworthy digital content—ensuring you know where content came from, who touched it, and whether it's real or AI-generated

5.4: First-Person Proof, Delegation, and Decentralized Trust Graphs

Drummond Reed, The First Person Project

The Problem: How Do You Prove You're a Real, Unique Person Online?

- Proving personhood (you're real, unique, and not a bot) is one of the oldest, toughest problems in digital identity
- Current solutions—especially biometric ones—have issues with privacy and centralization

- Vitalik Buterin favors social graph-based approaches (connections between real people), but notes they need stronger privacy

The Solution: Decentralized Identity + Personal Credentials

- Using decentralized identifiers (DIDs) and verifiable credentials, it's possible to build a private, secure, and scalable social graph
- The "First-Person Graph" lives only in individual wallets and agents, not on any centralized server
- Inspired by the PGP key-signing model, this approach enables people to form trust relationships peer-to-peer

Two Credential Types Power the Graph

1. **Personal Credentials**: Issued by trusted institutions (e.g., governments, schools, employers) asserting uniqueness per person, with privacy via zero-knowledge proofs
2. **Verified Relationship Credentials**: Issued between individuals (like digital handshakes), creating a web of trust

How It Works in Practice

- Alice and Bob exchange QR codes and DIDs using wallets
- They establish a private communication channel
- They then issue signed credentials to each other—verifying a trusted relationship without any intermediary

The First-Person Network

Each institution or community that issues personal or relational credentials becomes part of a decentralized trust ecosystem. These ecosystems can be connected to form a global web of trust, like a decentralized version of LinkedIn—but user-controlled.

What About AI?

The next frontier is ensuring not just human identity, but that AI agents act on behalf of real people. For example, "First-Person Agents"—AI tools humans can control and delegate power to, anchored in a living person's verified identity.

5.5: ITU's Work on AI Agent Security

Xiaoya Yang, ITU

ITU is formalizing global AI agent security standards to align with government regulation. Their multi-stage framework and focus on threat sources aim to build a technical backbone for AI safety at scale.

What is the International Telecommunication Union (ITU)?

ITU is an intergovernmental organization focused on global telecommunications standards. Unlike open-source groups like the Linux Foundation, ITU involves governments and regulators. Its goal is to bridge the gap between regulatory requirements and technical implementation.

AI Strategy Framework

ITU defines AI broadly, including: Machine learning, deep learning, and generative AI

ITU's efforts can be divided into three categories:

- Security of AI (threat modeling, attack mitigation)
- AI for Industry (finance, health, education, etc.)
- AI for Society (privacy, misinformation, fraud prevention)

AI Security Risk Types

- Endogenous risks: Risks from within AI systems (e.g., hallucination)
- Derived/external risks: Risks from attackers, users, tools, or agents themselves

AI Agent Security (Currently In Progress)

ITU is developing a standard called X.S-AIA (Security Requirements and Guidelines for Artificial Intelligence Agents).

AI agents are classified in stages:

- Perception & cognition
- planning
- memory
- action

For each stage, specific functions, threat sources, and security requirements are mapped.

Get Involved - Agentic AI

Subgroup Home Page	https://globaldigitalcollaboration.groups.io/g/GDC-AIForHumanity-AIAgent
Post	GDC-AIForHumanity-AIAgent@globaldigitalcollaboration.groups.io
Subscribe	GDC-AIForHumanity-AIAgent+subscribe@globaldigitalcollaboration.groups.io
Unsubscribe	GDC-AIForHumanity-AIAgent+unsubscribe@globaldigitalcollaboration.groups.io

Conclusion

Building AI with humanity in mind requires sustained global collaboration. The GDC provides a valuable platform for uniting diverse stakeholders in pursuit of this vision. The AI for Humanity Program will continue advancing this work while fostering dialogue through mailing lists, online forums, and in-person meetings throughout the year, with the aim of making a meaningful contribution to next year’s GDC.

Acknowledgements

The AI for Humanity Track was primarily organized through the leadership of Anni Lai, Co-Chair of Generative AI Commons; Board Director of LF AI & Data; Member of the Advisory Board, LF Europe; and Head of Open Source Operations, Futurewei. Appreciation is also extended to the presenters and panelists for their time and expertise: Anni Lai, Rob Geada, Mac Misiura, Sunu Engineer, Kacholiya, Gabriele Columbro, Yonghua Lin, Dong Sun, Sachiko Muto, Dylan Reim, Steven Tebbe, Wenjing Chu, Eric Drury, Scott Perry, Drummond Reed, and Xiaoya Yang. Special thanks are given to Daniel Goldscheider, Executive Director of the Open Wallet Foundation, for his role in founding and organizing GDC and for supporting the creation of the AI for Humanity Track.

About the author

Ethan Westfall was born and raised in the state of Idaho, USA. After graduating from high school as an accomplished student athlete he spent two years fulfilling a service assignment in Java, Indonesia. On a full-ride academic scholarship he then pursued undergraduate studies at Brigham Young University, graduating Summa Cum Laude with a degree in Business Strategy. Following this, he attended Harvard University, obtaining a master's degree in Learning Technology. The last few years he has worked in product and software engineering at one of the premier financial institutions in the United States. In addition to this day job, he enjoys pursuing a keen interest in AI-related endeavors, serving as secretary for an IEEE working group that aims to establish benchmarks for the development of AI.