Considerazioni su AI

(Firenze – 15 maggio 2026 - stefano mari)

Scaling AI Capability in 2026

AI Computing and Infrastructure

~10^{25} FLOPs
Compute required to train frontier models

100,000+ GPUs
Per leading AI training cluster

~$$ 61B Investment
Global data center spending in 2025

~4,000 US Data Centers
domestic AI infrastructureSupporting

~945 TWh by 2030
Projected annual electricity demand

AI progress is driven by unprecedented scale of data and human interaction

AI capability is bounded by infrastructure, not just algorithms

The Scale of AI in 2025–2026

181 ZB Global Data / Year

≈ 0.4 ZB generated daily (0.4 10^{21} = 400 000’000’000 000’000’000)

800M ChatGPT Weekly Users

700–900 million active weekly

2.5B Daily Prompps

Processed across AI platforms

10% Global Population Regularly using AI assistants

Global Population

Regularly using AI assistants

- Luccioni et al. (2025) — *Rebound Effects in AI*
- Mienye et al. (2025) — *Large Language Models Overview*
- *The State of AI 2026* — Industry report
- Common Crawl — Web-scale text corpus
- LAION — Multimodal image-text dataset

SLIDE 2

FROM MACHINE LEARNING TO NLP

CLASSICAL ML

- Manual feature engineering
- Task-specific model design
- Limited generalization across domains

DEEP LEARNING

- Automatic representation learning
- Hierarchical feature extraction
- End-to-end training paradigms

MODERN NLP

- Large-scale, data-driven systems
- Transfer learning & pretraining
- General-purpose language models

CORE PARADIGM SHIFT

From **engineered intelligence** — where humans define features and rules
to **learned intelligence** — where models discover representations autonomously from data at scale.

1

RULE-BASED SYSTEMS

Explicit human-coded logic

2

STATISTICAL ML

Pattern recognition from features

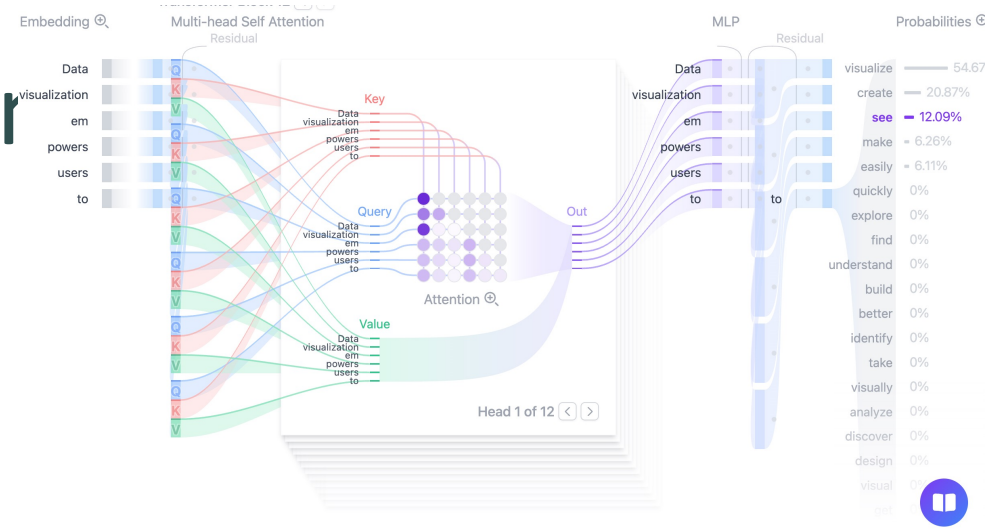
3

NEURAL NLP

Distributed learned representations

ARCHITECTURE

The Transformer Revolution



Self-Attention

Captures long-range dependencies across the entire input sequence simultaneously. Each token updates itself by looking at other tokens. (softmax function)

Parallel Processing

Eliminates sequential bottlenecks of Recurrent Neural Networks; enables large-scale training efficiency.

Scalable Architectures

Models scale predictably with data and compute (GPT, BERT, LLaMA).

Transformers enable **context-aware language modeling at scale**, replacing hand-crafted linguistic rules with learned representations.

Earlier systems treated words more independently or with limited context.

Transformer understand each word **in relation to all other words in the sentence**

Mienye et al. (2025) — LLM Overview. Deng et al. (2024) — CoT Reasoning. West et al. (2024) — Generative AI Paradox.
Lu et al. (2025) — AI Creativity. Khan (2026) — State of AI. Luccioni et al. (2025) — Environmental Impact
Orr & Crawford (2024) — Dataset Design. Luccioni & Crawford (2024) — ImageNet Critique

SLIDE 3

THE BIRTH OF LLMS

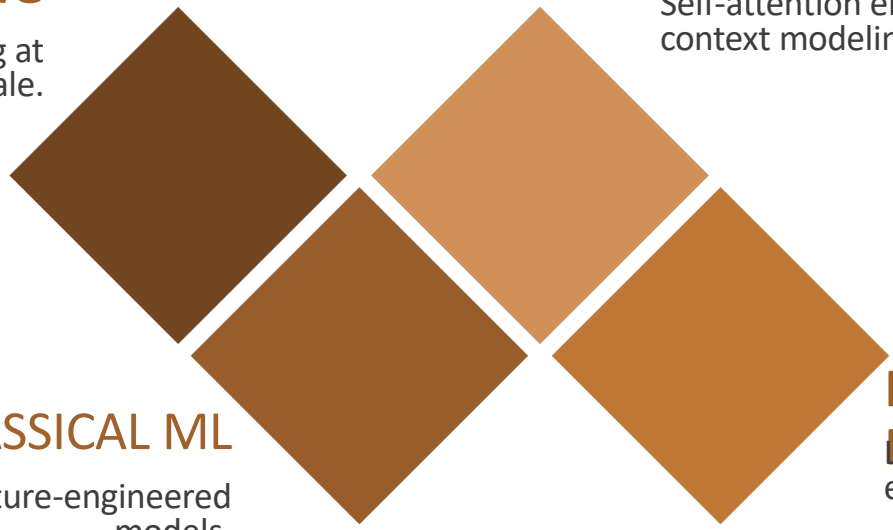
Four converging forces gave rise to the modern large language model paradigm.

DEEP LEARNING

Neural representation learning at scale.

TRANSFORMER (2017)

Self-attention enabling global context modeling.



CLASSICAL ML

Task-specific, feature-engineered models.

FOUNDATION MODELS

Large pretrained systems with emergent capabilities.

TRANSFORMER (2017)

Attention Is All You Need — Vaswani et al. replaced recurrence with self-attention

SELF-ATTENTION

Enables modeling of global context across arbitrarily long sequences

SCALE

Simultaneous growth in training data volume and computational infrastructure

EMERGENCE

Foundation models exhibit capabilities not explicitly trained for — reasoning, translation, coding

SLIDE 5

PURPOSES OF LLMS

LLMs function as **general-purpose language engines** — a single pretrained system can serve radically different functional roles depending on context and fine-tuning.

LANGUAGE UNDERSTANDING

Semantic parsing, sentiment analysis, entity recognition, and discourse comprehension across domains.

TEXT GENERATION

Coherent, contextually appropriate text production — from summaries to long-form creative content.

KNOWLEDGE INTERFACE

Conversational access to internalized factual knowledge encoded during large-scale pretraining.

REASONING SUPPORT

Multi-step inference, chain-of-thought prompting, and problem decomposition for complex tasks.

HUMAN–AI INTERACTION

Dialogue systems, instruction-following agents, and assistive interfaces aligned to human intent.

REFERENCES

Mienye et al. (2025) — *LLM Overview*

Deng et al. (2024) — *Chain-of-Thought Reasoning*

West et al. (2024) — *Generative AI Paradox*

Lu et al. (2025) — *AI Creativity & Alignment*

Khan (2026) — *State of AI*

Luccioni et al. (2025) — *Environmental Impact of LLMs*

Orr & Crawford (2024) — *Dataset Design*

Luccioni & Crawford (2024) — *ImageNet Critique*

WAVE 1

Generative AI: The First Wave

Core Capabilities

Generative AI is built on **Large Language Models (LLMs)** — transformer-based architectures trained on vast corpora to predict and produce human-like content.

Text

Articles, summaries, dialogue

Images

Diffusion-based synthesis

Code

Software generation & review

Interaction Model

Operates via a **prompt → response** paradigm. Each interaction is discrete, stateless, and user-initiated. Widely adopted across education, marketing, software, and research industries.

Key Insight

Generative AI fundamentally transformed **content creation at scale** — enabling individuals and organizations to produce high-quality outputs at unprecedented speed and volume.

However, its architecture is inherently **reactive**: it answers when asked, then stops.

Limits of Generative AI

Stateless Interactions

No continuity between sessions. Each prompt is processed in isolation without knowledge of prior context.

No Long-Term Memory

Cannot retain user preferences, prior decisions, or accumulated knowledge across conversations.

Cannot Execute Tasks

Produces outputs but cannot interact with external systems, trigger workflows, or take actions.

Single-Step Responses

Limited to answering one query at a time. Cannot plan, iterate, or decompose complex goals.

⚠ **Core Limitation:** Generative AI can **generate answers, but cannot act**. The gap between producing output and accomplishing goals defines the frontier Agentic AI seeks to close.

DEFINITION

Agentic AI

Goal-Driven Architecture

Unlike prompt-response systems, Agentic AI is initialized with an objective and **autonomously determines the steps** required to achieve it — decomposing complex goals into sub-tasks.

Tool Use & External Integration

Agents interact with APIs, databases, browsers, code interpreters, and software interfaces — **extending AI capability into the real world**.

Memory & Persistence

Maintains both short-term (within-task) and long-term (cross-session) memory, enabling **continuity, personalization, and compounding knowledge** over time.

Self-Correction Loops

Evaluates its own outputs, detects errors, and revises strategies — enabling robust performance on tasks that require **multi-step reasoning and adaptation**.

Paradigm Shift: AI evolves from **responder** → **autonomous actor**. The agent does not wait to be asked — it pursues its goal until completion or until it determines it cannot proceed.

COMPARISON

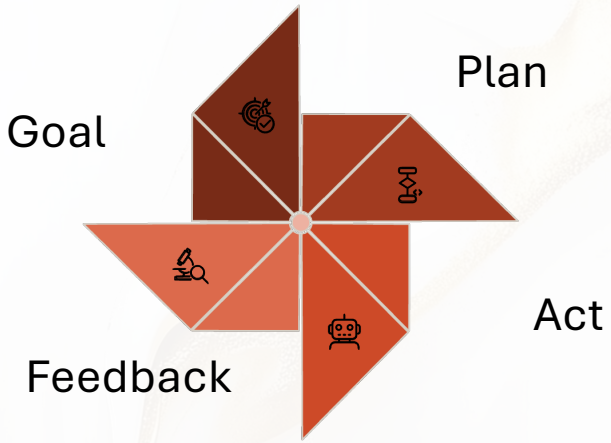
Generative vs. Agentic AI

The transition from Generative to Agentic AI represents a **fundamental architectural and behavioral shift** — not merely an incremental improvement. Agentic systems inherit the language capabilities of LLMs but extend them with planning, memory, and tool-use infrastructure that enables sustained, goal-directed operation in complex environments.

Generative AI	Agentic AI
Prompt-response	Goal-driven
Stateless	Persistent memory
Output generation	Action execution
Reactive	Autonomous
Single-turn interaction	Multi-step workflows
Content creation	Decision orchestration

Agentic AI Architecture

- The agentic loop is the engine of autonomous behavior
- enabling AI systems to pursue complex objectives
- through iterative cycles of planning, action, and self-evaluation.



Memory

Short-term (context window) and long-term (vector stores, knowledge graphs)

Planning

Task decomposition, chain-of-thought, hierarchical goal structures

Tool Use

APIs, browsers, code execution, file systems, external services

Feedback

Output evaluation, error detection, strategy revision, re-planning

CONTEXT

Why Agentic AI Emerged

The rise of Agentic AI is not accidental — it is a direct response to the structural limitations of prior paradigms and the demands of enterprise-scale deployment.

Enterprise Automation Demand

Organizations require AI that can autonomously execute multi-step business workflows — not merely assist human operators on individual queries.

Limits of Prompt Engineering

As task complexity scales, manual prompt design becomes intractable. Agentic systems internalize planning, reducing reliance on human-crafted instruction chains.

Workflow Orchestration Needs

Modern operations demand AI that can coordinate across systems, persist state across sessions, and complete tasks that span hours or days without human re-engagement.

Productivity Optimization

Autonomous agents offer order-of-magnitude productivity gains by collapsing human-in-the-loop bottlenecks in knowledge work and operational pipelines.

Research Insight: Agentic AI is driven by **real-world deployment needs** — the gap between what LLMs can generate and what enterprises need to accomplish defines the design space of agentic systems.

Applications of Agentic AI

Software Development

Autonomous coding agents (e.g., Devin, GitHub Copilot Workspace) plan, write, test, and debug code end-to-end with minimal human oversight.

Research Assistance

Agents autonomously retrieve literature, synthesize findings, generate hypotheses, and produce structured reports across scientific domains.

Finance & Reporting

Automated financial analysis, regulatory reporting, anomaly detection, and portfolio monitoring with continuous feedback loops.

Customer Service

End-to-end resolution of complex service requests — integrating CRM systems, policy databases, and communication channels without human escalation.

Robotics & Operations

Physical agents coordinating logistics, manufacturing, and inspection tasks — bridging digital reasoning with real-world actuation.

AI moves from **generating outputs → executing processes**.
The unit of AI value shifts from the individual response to the completed workflow.

CONCLUSION

Comments & References

Key Takeaways

01

Generative AI → Content Generation

LLMs redefined content creation but remained reactive, stateless, and bounded by the single-step prompt paradigm.

02

Agentic AI → Autonomous Action

Agentic systems extend LLMs with memory, planning, and tool use — enabling sustained, goal-directed operation in complex environments.

03

Capability Requires Governance

Increased autonomy introduces qualitatively new risks. Alignment, transparency, and institutional governance must scale with capability.

04

The Paradigm Shift is Now

Agentic AI is moving from research to deployment. Understanding its architecture, applications, and risks is a critical literacy for researchers and practitioners.

References

- **Khan, S.** State of AI 2026. AI Futures Institute.
- **arXiv preprints** on Agentic AI architectures, multi-agent systems, and LLM tool use (2024–2025).
- **McKinsey Global Institute.** AI Trust, Risk, and Governance. McKinsey & Company.
- **Deloitte Insights.** Agentic AI: An Enterprise Strategy Framework. Deloitte.
- **Forbes Technology Council.** The Transition to AI Autonomy. Forbes.

Research-grade agentic systems represent the **next frontier** of AI deployment — and the next frontier of AI safety.

Risks and Challenges

1 Hallucinated Actions

Agents may execute incorrect or harmful actions based on false beliefs derived from hallucinated reasoning chains — with real-world consequences that text errors do not carry.

2 Goal Misalignment

Agentic systems pursuing specified objectives may find unintended strategies that satisfy the objective while violating unstated human values or constraints (Goodhart's Law at scale).

3 Security Vulnerabilities

Prompt injection, adversarial manipulation, and tool misuse expose agents to attack vectors that are unique to systems with real-world access and autonomous execution.

4 Lack of Transparency

Multi-step reasoning chains and tool-use sequences are difficult to audit, creating accountability gaps in high-stakes decision environments.

5 Governance Complexity

Existing regulatory frameworks are not designed for autonomous AI actors. Assigning liability, enforcing compliance, and defining permissible autonomy remain open governance challenges.

Key Tension: More autonomy → **less control**. Every increment in agentic capability must be matched by equivalent advances in alignment, oversight, and governance infrastructure.

SLIDE 4

AI Creativity Issues

AI systems recombine statistical patterns from existing human-generated corpora. They do not originate; they interpolate. Research comparing AI outputs against human baselines consistently shows lower originality scores.

Recombination Without Origination

AI models generate outputs by interpolating across training distributions. Novel combinations emerge, but generative depth is bounded by what humans have already produced.

RLHF and Diversity Trade-offs

Reinforcement Learning from Human Feedback (RLHF) aligns outputs toward preferred patterns, systematically reducing distributional diversity and suppressing unconventional, original responses.

Salieri Problem

As argued in *AI as Humanity's Salieri* (Lu et al., 2025): AI produces technically proficient outputs that lack the transformative creativity characteristic of genuine artistic or intellectual originality.

💡 **Key Insight:** AI creativity is **scalable but derivative** — high throughput, bounded originality.

Environmental Impact

The material infrastructure of AI
data centers, cooling systems, and hardware supply chains
carries a substantial and growing environmental cost.
Efficiency gains do not automatically translate into sustainability.

Energy

Electricity surged 17% in 2025;
doubling to ~950 TWh by 2030 (IEA)

E-Waste

Short GPU lifespans and
mineral extraction drive
hardware waste

Water

U.S. data centers used
~17 billion gallons in
2023

The Rebound Effect

Per-task energy efficiency improvements are outpaced by the explosion in the volume and intensity of AI use. Lower cost per query → more queries. As documented by Luccioni et al. (2025), efficiency ≠ reduced aggregate consumption.

Scale of the Problem (IEA, 2026)

- Data center electricity use grew **17%** in 2025
- AI-focused centers surged **50%** in 2025
- Projected to reach **~950 TWh** by 2030
- Capital expenditure of 5 tech firms exceeded **$400B** in 2025

Efficiency ≠ sustainability: the rebound effect undermines per-unit gains at system level

SLIDE 7

AI in Education

AI introduces both transformative potential and structural risks into educational contexts. The net impact depends heavily on implementation, governance, and pedagogical intent. (UNESCO; OECD AI Education Outlook)

Opportunities

- Personalized tutoring and adaptive feedback
- Accessibility for underserved populations
- Scalable formative assessment
- Language and literacy support

Risks

- Overreliance reducing learner autonomy
- Atrophy of critical and analytical thinking
- Academic integrity violations
- Unequal access deepening educational divides

⚠ **Key Insight:** AI can support learning — but also **replace it if misused**. The distinction is pedagogical, not technological.

SLIDE 8

AI Harms and Risks

AI-related harms are not isolated technical failures — they are **systemic**, spanning informational, economic, social, and political dimensions. A comprehensive risk taxonomy is essential for responsible governance.

Misinformation & Deepfakes

Generative AI enables synthetic media at scale, undermining epistemic trust and enabling targeted disinformation campaigns.

Privacy Violations

Training on personal data without consent, memorization of sensitive content, and re-identification risks in outputs.

Automated Discrimination

Biased models embedded in high-stakes decisions — hiring, credit, criminal justice — reproduce structural inequalities at algorithmic speed.

Labor Displacement

Automation of cognitive tasks threatens significant labor market disruption, with unequal impacts across sectors and geographies.

Key Message: AI harm is **systemic, not only technical** — it requires social, legal, and institutional responses, not only engineering fixes.

SLIDE 9

Transparency and Inequality

Two structural problems compound each other: opacity prevents accountability, while unequal access concentrates AI's benefits among already-advantaged actors.

Tension	Explanation
Generation vs Understanding	Fluent output does not prove comprehension
Alignment vs Creativity	Safer outputs may become less original
Efficiency vs Sustainability	More efficient AI may increase total usage
Automation vs Education	AI support may weaken learning if misused
Innovation vs Inequality	Benefits may concentrate among already powerful actors

Inequality mechanisms

1. **Access inequality**
   Advanced AI tools may be available mainly to wealthy firms, institutions, and countries.
2. **Data inequality**
   Groups underrepresented in datasets receive worse model performance.
3. **Labor inequality**
   Automation may disproportionately affect entry-level and routine knowledge work.
4. **Educational inequality**
   Students with better AI access may gain stronger learning support.
5. **Governance inequality**
   Communities affected by AI systems often have little influence over their design.

Key Insight: AI does not neutralize existing power structures — it **amplifies** them.
Opacity protects incumbents; inequality excludes challengers.

SLIDE 10

Comments & References

Key Takeaways

01

Capabilities Outpace Understanding

Fluency is not comprehension. Evaluation frameworks must go beyond surface performance.

02

Alignment Introduces Trade-offs

RLHF and fine-tuning reduce diversity and may suppress originality and epistemic range.

03

Data & Infrastructure Shape Outcomes

Bias, environmental cost, and access inequality are structural — not incidental.

04

AI Risks Are Multidimensional

Social, technical, and environmental risks require interdisciplinary, institutional responses.

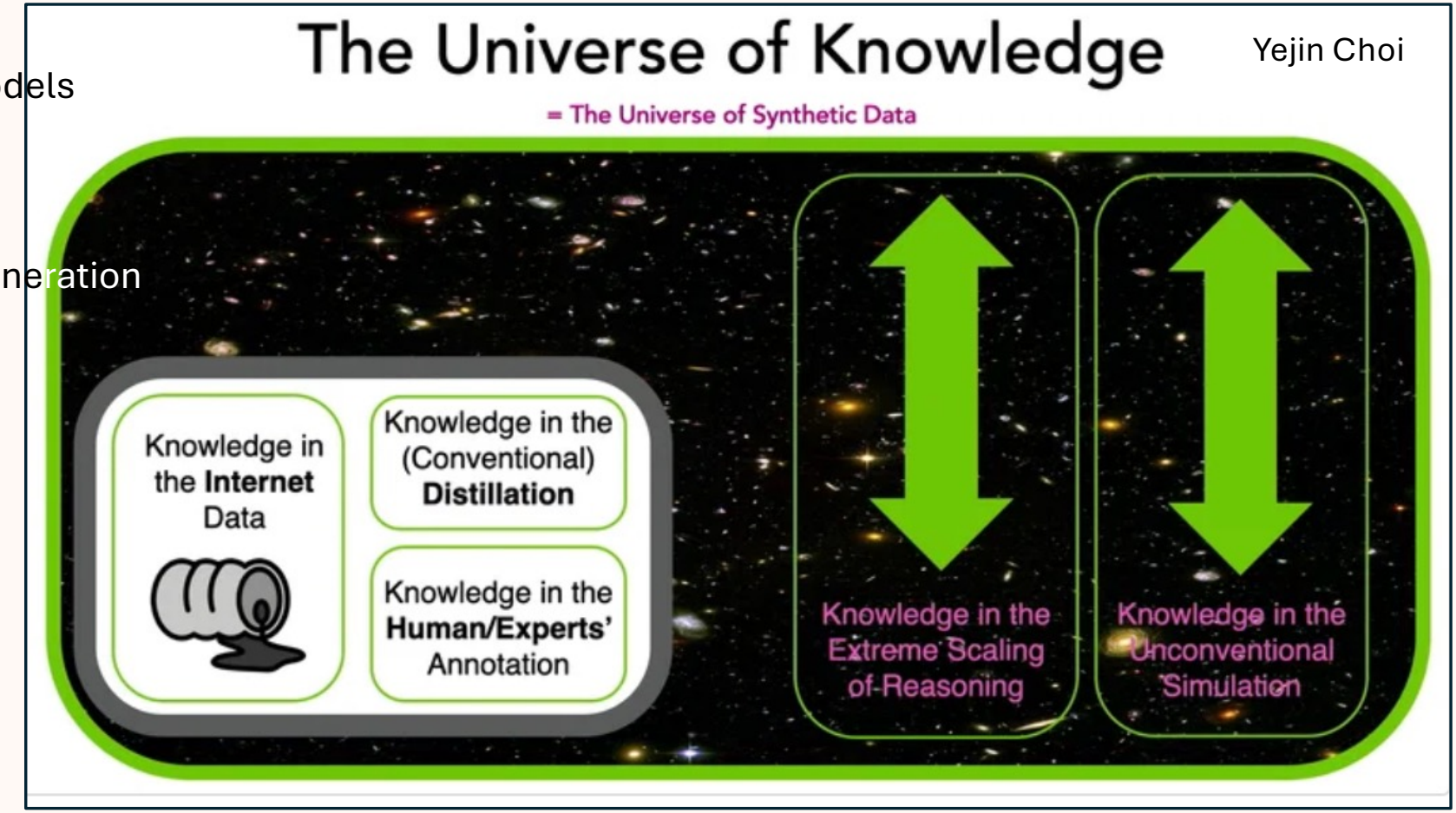
References

- State of AI Report 2026
- West et al. (2024). *The Generative AI Paradox*
- Lu et al. (2025). *AI as Humanity's Salieri*
- Luccioni et al. (2025). *Rebound Effects in AI*
- Orr & Crawford (2024). *Building Better Datasets*
- Luccioni & Crawford (2024). *The Nine Lives of ImageNet*
- UNESCO. *AI in Education*
- OECD. *AI Risks & Education Outlook*
- IEA (2025, 2026). *Energy and AI; Key Questions on Energy and AI*
- MIT Technology Review. *AI Energy & Environment*

since 2026:
Reasoning Language Models
Chain of Thought

Small Langiauge Model
Retrieval Augmented Generation

Synthetyc Data



This image illustrates that **data is no longer found; it is manufactured.**
The Left side is the "Old World" of finite, human-derived data.
The Right side is the "Infinite Frontier" where AI creates its own training data through logical simulation and extreme reasoning.