AI Toolkit — Grounded Link Extracts (Batch 5: AI Literacy & Model Limits)

Access date: January 27, 2026. Expanded grounded extracts for AI literacy and model limitations.

1. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?

URL: https://dl.acm.org/doi/10.1145/3442188.3445922

Source: FAccT Conference (Bender et al.) • **Date:** 2021

Key excerpt (≤25 words): “We refer to these systems as stochastic parrots.”

Why this matters: Foundational paper explaining why large language models reproduce patterns without understanding, including risks of bias, misinformation, and environmental cost.

AI-ingestible extract: Bender et al. argue that large language models generate fluent text by modeling statistical patterns in training data rather than grounded understanding, which can amplify bias, misinformation, and harmful stereotypes at scale.

2. GPT■4 Technical Report

URL: https://cdn.openai.com/papers/gpt-4.pdf

Source: OpenAI • **Date:** 2023

Key excerpt (≤25 words): “GPT■4 exhibits human■level performance on various professional benchmarks.”

Why this matters: Primary technical documentation describing capabilities and limitations of GPT■4, including hallucinations and reliability challenges.

AI-ingestible extract: The GPT■4 report documents improvements in reasoning and safety but notes persistent limitations including hallucinated facts, overconfidence, and sensitivity to prompt phrasing.

3. Why Language Models Hallucinate (OpenAI Blog)

URL: https://openai.com/research/why-language-models-hallucinate

Source: OpenAI • **Date:** 2023

Key excerpt (≤25 words): “Hallucinations occur when a model generates incorrect information that sounds plausible.”

Why this matters: Explains causes of hallucinations in generative AI systems and why they are difficult to eliminate completely.

AI-ingestible extract: OpenAI explains hallucinations arise from the probabilistic nature of language modeling, where the system predicts likely tokens rather than verifying facts, leading to confident but incorrect outputs.

4. The Myth of Artificial Intelligence

URL: https://www.nature.com/articles/d41586-021-02907-3

Source: Nature • **Date:** 2021

Key excerpt (≤25 words): “Today’s AI systems are not intelligent in the human sense.”

Why this matters: Editorial arguing that AI lacks true understanding and should be framed as advanced statistical tools rather than intelligent agents.

AI-ingestible extract: The article argues that current AI systems operate through pattern recognition and statistical correlation rather than reasoning or comprehension, and warns against anthropomorphic framing.

5. Attention Is All You Need

URL: https://arxiv.org/abs/1706.03762

Source: NeurIPS (Vaswani et al.) • **Date:** 2017

Key excerpt (≤25 words): “We propose a new simple network architecture, the Transformer…”

Why this matters: Foundational transformer paper explaining architecture behind modern LLMs, grounding technical explanations of how models process language.

AI-ingestible extract: Vaswani et al. introduce the Transformer architecture, which relies entirely on attention mechanisms to model relationships between tokens in a sequence without recurrent networks.

6. Model Cards for Model Reporting

URL: https://arxiv.org/abs/1810.03993

Source: Google Research • **Date:** 2019

Key excerpt (≤25 words): “Model cards are short documents accompanying trained machine learning models…”

Why this matters: Introduces the concept of model cards to document limitations, intended uses, and ethical considerations of AI systems.

AI-ingestible extract: Mitchell et al. propose model cards as standardized documentation describing performance, intended uses, limitations, and ethical considerations to improve transparency and accountability.

7. On the Opportunities and Risks of Foundation Models

URL: https://arxiv.org/abs/2108.07258

Source: Stanford Center for Research on Foundation Models • **Date:** 2021

Key excerpt (≤25 words): “Foundation models are trained on broad data at scale…”

Why this matters: Comprehensive report analyzing societal risks and benefits of large foundation models.

AI-ingestible extract: The Stanford report outlines both opportunities and systemic risks of foundation models, including bias propagation, environmental impact, misuse potential, and concentration of power.

8. AI Index Report

URL: https://aiindex.stanford.edu/report/

Source: Stanford HAI • **Date:** 2024

Key excerpt (≤25 words): “AI systems are becoming more capable but also more complex.”

Why this matters: Annual benchmark report tracking global AI progress, performance trends, and emerging risks.

AI-ingestible extract: The AI Index aggregates data on AI research, adoption, and performance benchmarks, highlighting rapid capability growth alongside rising concerns around safety, misuse, and governance.