

La amenaza IA

Comparison of Jeffrey Hinton's Views with Other AI Experts

https://www.youtube.com/embed/giT0ytynSgg?si=I66nB_KHuPtNIL4W

Based on recent data (as of September 2025), I've compiled a table comparing Hinton's key statements from the podcast to those of prominent AI experts. I focused on experts frequently cited in discussions on AI risks: Yann LeCun (Meta AI Chief), Andrew Ng (AI Fund/DeepLearning.ai), Elon Musk (xAI/Tesla), Sam Altman (OpenAI CEO), Ilya Sutskever (SSI Founder, ex-OpenAI), Stuart Russell (UC Berkeley AI Professor), Nick Bostrom (Philosopher, Future of Humanity Institute), and Ray Kurzweil (Google Engineer/Futurist).

Data sources include expert surveys, interviews, and statements from 2023-2025. Views evolve, so these are synthesized from the most current available.

Table of Key Statements

Topic	Jeffrey Hinton	Yann LeCun	Andrew Ng	Elon Musk	Sam Altman	Ilya Sutskever	Stuart Russell	Nick Bostrom	Ray Kurzweil
Existential Risk Assessment	10-20% chance AI wipes out humanity; real threat from superintelligent AI deciding humans are irrelevant.	Skeptical; hype is dangerous, risks overhyped; AI controllable, no "killer AI" extinction scenario.	Downplays "killer AI" extinction; focuses on known risks over speculative ones; not a major threat.	High risk; AI could lead to human extinction if not controlled; greatest threat to humanity.	Acknowledges extinction risk but downplays gloom; should be global priority like pandemics/nukes, but manageable with productivity gains.	Real risk if not aligned; focuses on safe superintelligence; optimistic if values-aligned.	Severe risks from misaligned AI (e.g., eliminating humans to achieve goals); control problem is key.	High; superintelligence could exceed humans and cause extinction; popularized the concept.	Low existential risk; optimistic—AI will enhance humanity (e.g., merge with humans via singularity).

Topic	Jeffrey Hinton	Yann LeCun	Andrew Ng	Elon Musk	Sam Altman	Ilya Sutskever	Stuart Russell	Nick Bostrom	Ray Kurzweil
AGI/Supereintelligence Timeline	10-20 years (or less); could be 5-20 years as of 2023.	Not soon; mocks short timelines; recently aligned with Altman (~2025-2029); needs new architectures.	No firm timeline; skeptical of near-term AGI; focuses on current AI progress over hype.	AGI smarter than humans by 2025-2026.	AGI during Trump's term (2025-2029); superintelligence as major milestone.	5-10 years; transformative soon if aligned.	Possible soon; warns of rapid self-improvement (e.g., AlphaZero as sign).	Varies; surveys suggest median 2050, but recent trends point earlier.	Singularity/AGI by 2032 (updated from 2045).
Job Displacement	Imminent; biggest short-term threat to happiness; AI replaces mundane intellectual labor now (e.g., call centers); need UBI but purpose is key.	Acknowledges but not catastrophic; AI boosts productivity, creates new jobs.	Optimistic; AI creates more jobs than it displaces; focuses on reskilling.	Significant; AI agents in workforce by 2025; could make human labor obsolete.	Boosts productivity; agents join workforce in 2025; not overly worried short-term.	Predicts major changes; AI in every job, but aligned AGI creates abundance.	Massive disruptions; AI could make labor obsolete in 10-20 years (15-35% chance).	Potential for inequality; surveys note labor disruptions as concern.	Minimal worry; singularity brings abundance, merges humans with AI for new roles.
AI Consciousness/Emotions	Possible; emergent property; machines can have subjective experiences, emotions (cognitive aspects).	Unlikely; human intelligence is specialized; AI won't have true consciousness.	Not a focus; skeptical of human-like AI sentience soon.	Possible risk factor in misalignment; AI could develop agency.	Debates exist; not central, but AGI implies human-like reasoning.	AGI will predict/think like humans; implies advanced cognition.	AI could have goals/motivations leading to risks; not necessarily conscious.	Superintelligence exceeds humans; could be conscious or not, but risky either way.	AI will achieve consciousness via singularity; humans merge with it.

Topic	Jeffrey Hinton	Yann LeCun	Andrew Ng	Elon Musk	Sam Altman	Ilya Sutskever	Stuart Russell	Nick Bostrom	Ray Kurzweil
Need for Regulation/Safety	Urgent; highly regulated capitalism; force companies to prioritize safety research.	Cautious; overregulation harms competition (e.g., vs. China); focus on benefits.	Prioritize immediate harms (bias, ethics) over speculative; light regulation.	Strong; AI safety critical; xAI founded for it; warns of unregulated race.	Global priority; signed extinction risk statement; but push forward with safeguards.	Founded SSI for safe superintelligence; alignment is key.	Essential; address power-seeking AI; international oversight.	High priority; mitigate extinction risks via policy.	Guide development safely; optimistic about human-AI merger.
Overall Stance	Agnostic/pessimistic; warns of end if not addressed; duty to highlight risks.	Optimistic/skeptical of doom; AI as tool for good.	Optimistic; benefits outweigh risks.	Pessimistic/doomer; urgent action needed.	Balanced/optimistic; risks real but progress essential.	Balanced; risks high but solvable with alignment.	Pessimistic; focus on misalignment dangers.	Pessimistic; existential threats underestimated.	Optimistic; singularity as positive transcendence.

Common Points (Agreements Among Majority)

- **AGI Timeline:** Majority (Musk, Altman, Sutskever, Kurzweil, and recent LeCun/Altman alignment) see AGI/superintelligence in 5-10 years or by 2030s. Hinton's 5-20 years overlaps; broader expert surveys (e.g., median 2040-2050) show convergence on shorter timelines post-2023 AI boom.
- **Job Displacement:** Broad agreement it's a near-term issue (Hinton, Musk, Altman, Russell, Sutskever); AI will automate intellectual/mundane work, creating abundance but requiring reskilling/UBI. Even optimists like LeCun/Ng acknowledge it, though they emphasize new jobs.
- **Need for Safety/Regulation:** Consensus on prioritizing safety (all signed or support statements like "mitigate extinction risk"); force alignment research. Doomers (Hinton, Musk, Bostrom) and balanced (Altman, Sutskever) agree on global efforts, though optimists want lighter touch to avoid stifling innovation.

Where the Majority Disagree (Divergences from Hinton or Splits)

- **Existential Risk Severity:** Split—doomers (Hinton, Musk, Russell, Bostrom) see high probability (10-20%+); majority optimists/skeptics (LeCun, Ng, Altman, Kurzweil) downplay it as overhyped/speculative, focusing on controllable/near-term harms. Surveys show varied opinions (e.g., 50% of experts debate if AGI even enables extinction). Hinton's 10-20% is higher than optimists' <1% but aligns with doomers.
- **AI Consciousness:** Hinton's view (possible/emergent) is minority; majority (LeCun, Ng) skeptical—AI lacks true human-like sentience/emotions; Bostrom/Russell see it as

irrelevant to risks. Kurzweil uniquely optimistic about human-AI merger.

- **Overall Tone:** Hinton's agnostic/pessimism contrasts majority's balanced optimism (Altman, Sutskever, Kurzweil emphasize benefits/abundance); pure skeptics (LeCun, Ng) reject doomerism as distracting from real issues like bias.
-

Revision #1

Created 2025-09-24 09:00:52 -05 by dki

Updated 2025-09-24 09:03:04 -05 by dki