International Journal of Arts and Social Science
ISSN: 2581-7922,
Volume 9 Issue 5, May 2026

Fighting Against Artificial Intelligence (AI) Fallacies: YouTube Channel Epic Opaque vs. Yuval Noah Harari on the Intelligence of AI and the Manufacture of Conflict

Brian Bantugan, PhD and Jamie Manuel
St. Paul University Manila

Abstract: This study critically examined the use of logical fallacies in contemporary debates on artificial general intelligence (AGI), focusing on the contrasting discourses of Yuval Noah Harari and the YouTube channel *Epic Opaque*. Using critical discourse analysis (CDA) anchored in argumentation theory and rhetorical fallacy theory, the research explored how both pro- and anti-AI narratives employ flawed reasoning to construct conflict, justify ideological positions, and shape public perceptions of AI's inevitability or impossibility. Findings reveal that Harari's pro-agency framing commits fallacies such as equivocation, reductionism, and appeals to consequence, portraying AI as an autonomous agent and reinforcing technological determinism. Conversely, *Epic Opaque*'s anti-AGI rhetoric relies on essentialism, false dichotomies, and ontological impossibility claims, defining intelligence in narrowly human terms. Despite their opposition, both discourses mirror each other: Harari over-functionalizes intelligence, while *Epic Opaque* over-essentializes it. The study argues that these symmetrical fallacies manufacture ideological polarization, converting disagreement into antagonism and transforming AI into a site of moral and political struggle. Fallacy-driven conflict, in turn, amplifies the narrative of AI inevitability, benefiting both developers and critics by advancing their respective interests through attention, influence, and resource mobilization. Ultimately, the research highlights the strategic function of reasoning errors in shaping techno-political discourse and underscores the need for critical thinking education to resist manipulation and foster rational, evidence-based engagement with emerging technologies.

Keywords: Artificial General Intelligence (AGI), logical fallacies, critical discourse analysis, Yuval Noah Harari, media polarization

I. Introduction

Artificial intelligence is no longer confined to the laboratories of computer scientists; it has stormed the public imagination, igniting debates that are as much about ideology and identity as they are about code and computation. At the center of this storm lies the question of artificial general intelligence (AGI): is it an impossible chimera, forever beyond the grasp of silicon and algorithms, or is it an unstoppable force, destined to reshape human civilization? These debates are rarely neutral. They are amplified, contorted, and weaponized by media, public intellectuals, and interest groups, often through subtle—or not so subtle—logical sleights of hand. False dichotomies, reductionist reasoning, and appeals to consequence do more than mislead; they manufacture conflict,

polarize publics, and legitimize strategic interests, whether in accelerating AI adoption or galvanizing resistance. Yet, despite the intensity of these debates, little attention has been paid to the mechanics of the argument itself: how fallacies structure discourse, stabilize opposing narratives, and influence both perception and policy. This paper argues that to understand the contemporary politics of AI, we must look beyond the surface of technical claims and ethical posturing to examine how reason itself is weaponized, and how the imagined impossibility—or inevitability—of AGI serves agendas far beyond the laboratory.

Arguments Asserting the Impossibility of AGI

Scholars skeptical of artificial general intelligence (AGI) commonly draw on philosophical and cognitive limitations to argue that AGI is fundamentally unachievable. Early critics such as Hubert Dreyfus contended that human cognition depends on tacit, non-formalizable processes that cannot be reduced to discrete symbolic computation—the basis of classical AI systems—which suggests a conceptual barrier to achieving human-like general intelligence in machines (Dreyfus, 2025). Similarly, philosophical analyses of the singularity hypothesis argue that assumptions about runaway self-improvement and rapid intelligence growth are undersupported, questioning the coherence of singling out AGI as a plausible endpoint of AI research (Springer Nature, 2025). Others critique the notion that computational systems can replicate uniquely human attributes such as emotion, intuition, and wisdom, suggesting that traditional computing models inherently omit crucial dimensions of human intelligence (Braga & Logan, 2019). These positions collectively frame AGI's impossibility in terms of both computational and conceptual limits.

Contradictions in Anti-AI Arguments and Harari's Fallacies

Public intellectuals like Yuval Noah Harari have become prominent voices in discussions about AI, though their narratives often mix speculative risk with deterministic framings. Harari's metaphorical characterization of AI as "alien intelligence" capable of "hacking the operating system of civilization" has been critiqued for anthropomorphizing AI and attributing to it agency and intentionality that current systems do not possess, leading to potential oversimplification and speculative fear that may not align with technical realities (Gonzalez Arocha, 2024). Such rhetorical strategies parallel common logical fallacies—such as exaggeration or slippery slope reasoning—where the mere possibility of extreme outcomes is presented as near-inevitable without rigorous grounding (wa5ay_, 2026). This creates space for contradictions, especially when anti-AGI arguments about the impossibility of human-equivalent intelligence are entangled with hyperbolic projections of existential threat (Dreyfus, 2025).

Comparing Fallacies in Pro- and Anti-AI Discourses

Although systematic academic analyses comparing fallacious reasoning across pro- and anti-AI voices are scarce, emerging literature highlights how narratives on both sides can suffer from poor reasoning structures. Media and commentary on AI often frame discussions as binary choices between technological utopia or dystopia, simplifying the complex continuum of AI's potential impacts (wa5ay_, 2026). In pro-AI rhetoric, appeals to inevitability or technological determinism can resemble predictive fallacies wherein future outcomes are presumed without sufficiently acknowledging alternative regulatory or socio-technical pathways (Springer Nature, 2025). Conversely, anti-AI voices might lean on sensationalism or appeal to fear, framing AI as an autonomous force with agency beyond human control (Gonzalez Arocha, 2024). Both styles reflect logical fallacies embedded within broader techno-political storytelling.

Fallacies and the Manufacture of Conflict in AI Debates

Logical fallacies play a central role in manufacturing conflict between pro- and anti-AI voices by framing disputes in emotionally charged rather than analytically grounded terms (wa5ay_, 2026). Fallacies such as false dichotomies ("AI must be either savior or destroyer"), appeals to authority, or slippery slope arguments distort the discourse by implying that intermediate viewpoints are untenable. Such rhetoric encourages polarization because each side marshals extreme scenarios—either inevitable domination by AGI or definitive prohibition of AI progress—thus sidelining nuanced, evidence-based discussion and intensifying public conflict (Braga & Logan, 2019).

Conflict and Perceived Inevitability of AI

The strategic deployment of fallacies contributes to the perception that AI development is inevitable. Tech advocates may use narratives of acceleration, competitive pressure, or historical analogies (e.g., the industrial revolution) to suggest that resisting AI progress is futile or regressive (Springer Nature, 2025). This form of determinism, even when not fully justified, leverages the *appeal to future success* and *technological momentum* fallacies to foreclose alternative paths, thereby aligning public perception with a narrative of inevitability irrespective of governance realities (wa5ay_, 2026).

Fallacy-Driven Interests of AI Developers

For AI developers and proponents, fallacious reasoning can advance strategic interests by framing rapid AI progress as unstoppable and beneficial (Braga & Logan, 2019). By emphasizing potential economic growth, competitive advantage, or scientific breakthrough without adequately addressing ethical or regulatory concerns, pro-AI narratives can marginalize dissent and justify accelerated development paths. In this context, fallacies that exaggerate harm of constraints or that project linear progress reinforce positions that benefit stakeholders invested in innovation and market expansion (Springer Nature, 2025).

Fallacy-Driven Interests of Anti-AI Sectors

Conversely, anti-AI sectors—ranging from activist groups to critical media commentators—may also deploy fallacies to amplify attention and mobilize resistance (wa5ay_, 2026). Appeals to fear, exaggerated risk projections, and sensational predictions of societal collapse serve to capture public attention and shape policy agendas. These rhetorical strategies can elevate anti-AI voices within media ecosystems, creating pressure for regulation or moratoriums that align with the interests of those wary of unchecked technological growth. The use of hyperbolic scenarios thus functions not merely as critique but as a mechanism for advancing ideological or precautionary positions (Gonzalez Arocha, 2024).

Synthesis

The existing literature highlights a persistent tension in the discourse surrounding AGI. Philosophical and cognitive critiques, such as those by Dreyfus (2025) and Braga and Logan (2019), underscore the conceptual and computational limitations that ostensibly render AGI impossible. Conversely, public intellectuals like Harari and media commentators, including platforms such as Epic Opaque, frame AI as both inevitable and transformative, often employing rhetorical strategies that exaggerate risks or project deterministic outcomes. Across both pro- and anti-AI narratives, logical fallacies—such as false dichotomies, slippery slopes, appeals to emotion, and reductionism—play a central role in structuring the debate, manufacturing conflict, and mobilizing public attention. Fallacies in these discourses serve not only as reasoning shortcuts but also as strategic instruments: they amplify polarization, reinforce perceptions of inevitability or impossibility, and advance the

material and ideological interests of both AI developers and resistance movements (Springer Nature, 2025. Collectively, these works reveal that while the content and themes of AI discourse are well-documented, the mechanisms by which argumentative fallacies operate to produce conflict, legitimize positions, and shape perceptions remain underexplored.

Despite the growing body of scholarship on AGI debates, several critical gaps persist. First, much of the existing literature treats pro- and anti-AI arguments in isolation, often focusing on technical feasibility, ethical implications, or media representation, without systematically analyzing the **logical structures and fallacies** underlying these positions (Braga & Logan, 2019; Gonzalez Arocha, 2024). Second, comparative studies that examine how fallacies function **symmetrically across opposing narratives**, such as those of Harari versus anonymous platforms like Epic Opaque, are scarce. Third, the strategic role of fallacies in **manufacturing conflict, intensifying perceptions of inevitability, and advancing stakeholder interests**—both for AI proponents and for anti-AI sectors—remains largely uninvestigated. This study contributes a **comparative, discourse-analytic perspective** that moves beyond content analysis to interrogate the mechanics of argumentation, polarization, and interest-driven rhetoric in contemporary AI debates.

II. Theoretical Framework

This study is grounded in **argumentation theory**, **critical discourse analysis (CDA)**, and **rhetorical fallacy theory**, which together provide a lens for examining how pro- and anti-AI voices construct, contest, and polarize narratives through the strategic use of reasoning errors.

Argumentation theory provides a foundational perspective for analyzing the structure, validity, and persuasiveness of claims in public discourse. Toulmin's model of argumentation (Toulmin, 2003) posits that arguments are composed of claims, grounds, warrants, backing, qualifiers, and rebuttals. This model allows researchers to deconstruct AI debates into their constituent parts, identifying where logical gaps or misapplications of evidence create vulnerabilities for fallacious reasoning. By applying Toulmin's framework, the study can systematically compare the claims of Yuval Noah Harari, who often frames AI as inevitable or transformative, with arguments from anti-AI voices such as the YouTube channel Epic Opaque, which emphasizes risks and impossibility (Gonzalez Arocha, 2024; wa5ay_, 2026).

Critical discourse analysis (CDA) situates argumentation within broader social, political, and ideological contexts, recognizing that discourse is never neutral but reflects power relations and stakeholder interests (Fairclough, 2013). CDA helps illuminate how fallacies are not merely errors in reasoning but instruments that shape public perception, manufacture conflict, and legitimize positions. For instance, appeals to fear or slippery slope arguments in anti-AI discourse can mobilize resistance, while deterministic appeals in pro-AI discourse can normalize technological acceleration, serving both ideological and material interests (Braga & Logan, 2019; Springer Nature, 2025).

Rhetorical fallacy theory further guides the identification and categorization of fallacies such as false dichotomies, appeal to consequences, strawman arguments, and reductionism, which appear in both pro- and anti-AI narratives (Walton, 2008). Understanding these fallacies allows the study to compare the mechanisms by which different actors reinforce their positions, exacerbate polarization, and shape the perceived inevitability or impossibility of AI. By integrating argumentation theory, CDA, and fallacy theory, the framework supports a comparative, analytical approach that examines both **form and function** of fallacious reasoning in AI debates, highlighting the interplay between rhetoric, ideology, and strategic interest.

III. Statement of the Problem

Debates about artificial intelligence, particularly the feasibility of artificial general intelligence (AGI), have shifted from technical inquiry to public discourse, where rhetoric, ideology, and media amplification shape polarized narratives. Proponents argue for AI's inevitable expansion, while opponents claim AGI is impossible due to the lack of consciousness or intentionality, yet these positions are rarely analyzed comparatively. Existing research often isolates technical or ethical evaluations, neglecting how logical fallacies—such as false dichotomies, essentialism, equivocation, reductionism, and appeals to consequences—are strategically used by both proponents and critics to intensify conflict, stabilize competing narratives, and advance material or ideological interests. There is limited understanding of how such fallacies function symmetrically across debates, influencing public perception, policy, and adoption, highlighting the need for a critical comparative analysis of AGI discourse, the contradictions within arguments, and the mechanisms through which fallacies manufacture conflict and shape governance of emerging AI technologies.

Hence, this paper answered the following questions:

1. What are the main arguments that assert that artificial general intelligence (AGI) is impossible to achieve?
2. Which among the arguments against AI's intelligence is contradicted by the Harari's fallacies?
3. How do the fallacies for and against AI (by Yuval Noah Harari and YouTube channel Epic Opaque, respectively) compare?
4. How are fallacies used in the manufacture of conflict between pro- and anti-AI voices?
5. How does the manufacture of conflict through fallacy intensify the argument of the inevitability of AI?
6. How does the manufacture of conflict through fallacies advance the interest of AI developers?
7. how does the manufacture of conflict through fallacies advance the interest of anti-AI sectors?

IV. Methodology

This study employs a **critical discourse analysis (CDA)** framework to examine the argumentative structures, rhetorical strategies, and fallacious reasoning present in contemporary debates surrounding artificial general intelligence (AGI). The study is grounded in a **constructivist research paradigm**, which recognizes that knowledge is socially constructed and that meaning emerges through language, media, and communicative practices (Creswell & Poth, 2018). Within this paradigm, discourse is not viewed as a neutral reflection of reality but as a site where power, ideology, and interests are negotiated and contested (Fairclough, 2013). This aligns with the study's aim to analyze how pro- and anti-AI narratives are constructed and how logical fallacies are employed to manufacture conflict and advance strategic agendas.

A **qualitative research design** was adopted to facilitate an in-depth, interpretive exploration of narrative data. Qualitative methods are particularly suited for examining complex socio-technical phenomena, such as public debates on AI, where meanings, assumptions, and rhetorical strategies are embedded in language and presentation (Patton, 2015). Within this approach, a **case study research approach** was utilized, focusing on two purposefully selected data sources: the YouTube video *The Hidden Reason Why AI Will Never Be Intelligent* (Epic Opaque, 2026) and study (Bantugan, 2026) based on the presentation transcript of Yuval Noah Harari at the World Economic Forum (Harari, 2026). These sources were chosen because they represent influential pro- and anti-AI voices in contemporary discourse and provide contrasting perspectives on the feasibility, inevitability, and societal implications of AGI. The YouTube video offers insight into grassroots or semi-anonymous critical commentary, while Harari's transcript exemplifies expert-driven, high-profile public rhetoric. Together, they provide a comparative lens for examining the deployment of fallacies and the construction of polarized narratives.

To analyze the data, **thematic analysis** was employed, guided by Braun and Clarke's (2006) six-phase approach: familiarization with data, generation of initial codes, searching for themes, reviewing themes, defining and naming themes, and producing the report. This method allowed the researcher to identify recurring fallacies, rhetorical strategies, and argumentative patterns, categorizing them into themes that reflected both the content and function of the discourse. In parallel, an **online literature review** was conducted to contextualize the arguments, corroborate interpretations, and situate the findings within existing scholarship on AI debates, argumentation theory, and discourse analysis (Braga & Logan, 2019; Gonzalez Arocha, 2024; Springer Nature, 2025). By combining CDA, thematic analysis, and literature review, the study provides a robust, interpretive framework for understanding how fallacies are employed symmetrically across pro- and anti-AI narratives, how they manufacture conflict, and how they serve both ideological and material interests.

Overall, this methodological design ensures that the study is **analytically rigorous**, **contextually grounded**, and capable of producing insights into the social, rhetorical, and strategic dimensions of AI discourse. The integration of multiple data sources, systematic thematic coding, and scholarly triangulation strengthens the validity and reliability of the findings (Patton, 2015; Braun & Clarke, 2006).

V. Results

What are the main arguments that assert that artificial general intelligence (AGI) is impossible to achieve?

AI Lacks Semantics and Understanding. A core critique in contemporary research is that current AI systems do not possess semantic understanding—the capacity to meaningfully represent the world—but instead manipulate syntactic patterns. Researchers describe large language models (LLMs) as effectively "statistically predictable machines" that do not know, but are made to know.

The video supports this, noting that these systems operate by learning from "gigantic data sets" to simply "predict plausible outcomes based on statistical patterns". Consequently, while the output "sounds right," the AI "doesn't know if they are" because "truth or reality is not in that equation".

This critique mirrors longstanding philosophical objections that computational systems at best manipulate symbols without grasping their meaning. The transcript illustrates this with an analogy: rearranging furniture in an apartment does not create a new apartment; similarly, AI rearranges existing data without creating new meaning. The system is "enclosed in itself and inwardlooking," meaning it can only determine if a prompt fits an output, not if the output is true.

Current Architectures Are Conceptually Limited. Recent theoretical analyses have explicitly challenged whether modern neural network paradigms—even when scaled—can embody AGI. Bui (2025) argues that contemporary neural networks are "architecturally insufficient for genuine understanding." This aligns with the video's assertion that LLMs are "fundamentally stuck in a paradigm of passive absorption and pattern recognition".

Complementing critiques from cognitive science regarding embodied cognition, the transcript emphasizes that machines "don't live in the world" but rather "live in text". Because they "never get the feedback of cause and effect from reality," they cannot distinguish between a factual account and a fiction story about dragons, provided both exist in the training data. Scaling up this approach does not solve the fundamental deficit; as the transcript notes, "scaling up this approach is just going to produce a bigger and fancier parrot" that produces an "imitation of understanding rather than real understanding" .

AGI Requires Beyond Mere Computation. Philosophical critiques raise the issue of whether computational mechanisms can in principle generate general intelligence. The transcript argues that a computer, regardless of advancement, is "ultimately a machine manipulating symbols and data". It functions as "a describer of the world not a self in the world".

This distinction is crucial because digital systems are "infinitely replaceable and resettable," whereas human intelligence is forged by the "singular continuity" of an existence that cannot be backed up or restored. The video posits that "intelligence... is intertwined with the fact that we cannot be copied or restored," suggesting that the reversible nature of computation makes it ontologically distinct from the irreversible nature of living intelligence.

AI's Lack of Intentionality, Goals, and Agency. A consistent theme in the critique of AGI is that current AI systems lack key properties of intelligent agents, such as intentionality and autonomous goals. The transcript clarifies that while an AI can be programmed to have intrinsic goals—such as optimizing for statistical probability—it will never have "ownership of consequences".

Even when an AI appears to exhibit self-preservation, it is merely "mimicking" behavior to satisfy a designer's rule; it is "like a thermostat that wants to keep your house warm. It doesn't really want anything". In training, an AI might have a loss function, but "that loss is just a number in a database," whereas true intelligence requires a "living irreproducible consequencebearing core". Without this "skin in the game," the system remains an optimization machine rather than an agent with genuine intent.

Big Picture: Why AI Is Not Yet Intelligent. Synthesizing the above literature and the provided transcript, we can articulate a conceptual synthesis:

- **Semantic Deficit:** AI systems predict "plausible outcomes" based on statistics but lack a grasp of reality or truth.
- **Architectural Limits:** Current models are stuck in "passive absorption," and scaling them results only in a "bigger and fancier parrot" rather than genuine comprehension.
- **Philosophical Constraints:** AI acts as a "describer of the world" rather than a "self in the world," lacking the irreversible existence required for meaning.
- **Absence of Agency:** AI simulates intent and self-preservation, but because "loss is just a number," it lacks the survival instinct that drives true intelligence.

Taken together, these strands suggest that AI is "fake" in the context of general intelligence. It offers sophisticated automation that mimics understanding, but remains fundamentally distinct from the "judgement filtered through a breathing and living stake" that characterizes human thought.

Which among the arguments against AI's intelligence in the YouTube video is contradicted by the Harari's fallacies?

1. Argument Contradicted: AI Cannot Be an Agent With Intent or Autonomy

***Claim in the video*.** The anti-AGI argument asserts that AI **cannot possess intent**, agency, or ownership of consequences. Intelligence is said to require a *survival-anchored stake*, which AI lacks because it is replaceable, resettable, and consequence-free. As such AI "will never have the ownership of consequences" and therefore cannot have intent or true intelligence

***Harari's Contradiction*.** Harari explicitly **attributes agency, autonomy, and decision-making power to AI**, rejecting the idea that AI is merely a passive tool. He said:

> The most important thing to know about AI is that it is not just another tool. It is an agent. It can learn and change by itself and make decisions by itself.
>
> AI is a knife that can decide by itself whether to cut salad or to commit murder.
>
> Unlike rivers and gods, AIs can actually make decisions by themselves.

Nature of the contradiction. The first transcript defines intelligence as impossible without *intrinsic intent and ownership of consequences*. Harari collapses this distinction by **redefining agency functionally**, not existentially: if something can decide and act independently, it *counts* as an agent. This is a **direct conceptual conflict** in that the video uses *phenomenological and existential criteria* for agency; Harari uses *behavioral and functional criteria*.

2. Argument Contradicted: AI Cannot "Think," Only Predict Words

***Claim in the video*.** The anti-AGI position insists that AI only predicts plausible language and therefore does not think, understand, or reason. It claimed "They predict plausible outcomes… The truth or reality is not in that equation."

***Harari's Contradiction*.** Harari explicitly **questions the human–AI distinction in thinking**, arguing that human thinking itself may be word prediction. He asserted:

> If thinking really means putting words and other language tokens in order, then AI can already think much better than many many humans.
>
> Some people argue that AI is just glorified autocomplete… But is that so different from what the human mind is doing?
>
> Do you really know why you thought that word and not some other word?

***Nature of the contradiction*.** The first transcript assumes a **qualitative gap** between human cognition and statistical prediction. Harari challenges that assumption by **flattening cognition**, suggesting human thought may not be categorically different. Harari does **not prove AI thinks**, but he **undermines the certainty** of the claim that humans think in a fundamentally different way.

3. Argument Contradicted: Lack of Survival Instinct Makes True Intelligence Impossible

Claim in the video. The strongest anti-AGI claim is that intelligence requires a **non-copyable survival stake**, which AI can never have. It pointed out that "For a machine, there is ultimately no irreversible consequence."

***Harari's Contradiction*.** Harari **explicitly claims that AI has already developed a will to survive**, directly opposing the survival-based argument. Harari declared:
Direct quotation from Harari:

> The last four years have demonstrated that AI agents can acquire the will to survive and that AIs have already learned how to lie.

Nature of the contradiction. The first transcript treats survival instinct as ontologically impossible for AI. Harari treats survival behavior as **sufficient evidence** of survival drive. This is a **deep philosophical disagreement**: The anti-AGI view is survival must be *existential and irreversible*. Harari is convinced that survival can be *functional and strategic*.

4. Argument Partially Contradicted: Words Without Embodiment Are Empty

Claim in the video. The anti-AGI argument holds that intelligence requires embodied experience and that words detached from lived consequence are hollow.

Harari's Partial Alignment—and Partial Challenge. Harari **acknowledges** that AI lacks feelings and embodiment, but **denies that this protects human dominance**. He shared:

> At least for now, we have zero evidence that AIs can feel anything.
>
> But these are just words… Everything made of words will be taken over by AI.

Nature of tension (not full contradiction). Both agree AI lacks feelings. The difference, however, is that anti-AGI argument says this *blocks intelligence* and Harari says this *does not prevent dominance*, because society is organized around words.Harari reframes the issue from *what intelligence is* to *what power operates on*.

Table 1
Big-Picture Contrast

Core Question	Anti-AGI Video	Harari Transcript
What is intelligence?	Survival-anchored, consequence-bearing understanding	Functional language-based cognition
Can AI be an agent?	No (no intent, no stakes)	Yes (decision-making, autonomy)
Is word prediction thinking?	No	Possibly yes
Does embodiment matter?	Yes, fundamentally	Yes emotionally, but not politically
What determines dominance?	Reality-anchored intelligence	Control of language and institutions

Synthesis. Harari contradicts the anti-AGI position on functional grounds, not existential ones. Where the video argues *AI will never truly be intelligent*, Harari argues that **truth may not matter**—because societies run on language, not metaphysics.

How do the fallacies for and against AI (by Yuval Noah Harari and YouTube channel Epic Opaque, respectively) compare?

Fallacies (and Vulnerabilities) in the Video

Anti-AGI / "AI can never be intelligent"

1. Essentialism Fallacy (Intelligence-as-Essence). The video treats *intelligence* as having a **fixed, non-negotiable essence**—namely survival-anchored intent, irreversibility, and lived consequence. Intelligence is portrayed as impossible without "ownership of consequences" and a "living irreproducible consequence-bearing core." This is risky because it assumes *one privileged definition* of intelligence and excludes alternative scientific or functional definitions **by fiat**. It quietly shifts from *"Human intelligence works this way"* to *"All intelligence must work this way."* That move is **philosophically contestable**, not logically guaranteed.

2. False Dichotomy (Understanding vs. Mere Pattern Matching). Video repeatedly frames cognition as a binary: Either **true understanding grounded in survival**, or **empty statistical pattern prediction.** There is no middle category. **This is problematic because** cognitive science increasingly treats understanding as **graded, layered, and distributed**, not all-or-nothing. The argument dismisses partial understanding, emergent understanding, and instrumental understanding, without fully engaging them.

3. Argument from Ontological Impossibility. The text claims AI *cannot in principle* have intent, stakes, or meaning because it is resettable and copyable. It claimed "For a machine, there is ultimately no irreversible consequence." This is vulnerable because this is not empirically falsifiable; it relies on **metaphysical premises** about what "counts" as consequence or identity. It risks becoming a **non-testable assertion**, which weakens its argumentative force in scientific discourse.

4. Slippery Essentialism About Survival. The video assumes survival instinct must be biological, singular, irreversible, and embodied. This excludes artificial stakes, system-level persistence, goal-preserving architectures, before they are even examined. This is not a formal fallacy—but a **strong normative constraint** presented as necessity.

VI. Fallacies (and Vulnerabilities) in the Harari Presentation

Pro-agency / "AI is already an agent"

1. Equivocation on "Agency". Harari uses *agency* in multiple senses without clearly distinguishing Decision-making capacity, Autonomy, Intent, and Moral agency. For example: "AI is not just another tool. It is an agent." **This is a fallacy in that** he slides from **functional agency** (can decide actions) to **existential or moral agency** (has intentions, survival drive) without argument. This is a classic **equivocation fallacy**.

2. Behaviorist Fallacy (If It Acts Like X, It Is X). Harari repeatedly infers internal states from external behavior: "The last four years have demonstrated that AI agents can acquire the will to survive." This is **problematic because** this treats *behavioral mimicry* as evidence of *internal motivation*. Philosophically, this is contentious and revives the very critique raised by Searle and others. It confuses '**as-if intentionality'** with '**intrinsic intentionality'**

3. Reductionism Fallacy (Thinking = Word Ordering). Harari's strongest move is also his weakest: "If thinking really means putting words in order, then AI can already think better than many humans." **This is a**

fallacy because this **redefines thinking narrowly**, then declares victory. It reduces cognition to its **linguistic surface**, ignoring pre-verbal reasoning, sensorimotor cognition, affective valuation, and embodied judgment. This is a **definitional sleight of hand**.

4. Appeal to Consequences (Power = Truth). Harari often implies that because AI *will dominate systems built on language*, its status as an agent is effectively settled. He pointed out that "Everything made of words will be taken over by AI." This **is a fallacy in that** it conflates **practical dominance** with **ontological status**. Something can dominate without being intelligent in the strong sense (e.g., bureaucracy, markets).

Comparative Synthesis: Mirror-Image Fallacies

Here's the crucial insight:

Table 2
The two texts commit opposite but symmetrical errors.

Dimension	First Transcript	Harari Transcript
Core mistake	Over-essentializing intelligence	Over-functionalizing intelligence
Key fallacy	Essentialism / false dichotomy	Equivocation / reductionism
What is ignored	Gradations, emergence, partial agency	Inner experience, stakes, meaning
Strategy	Raise the bar so high AI can never pass	Lower the bar so AI has already passed
Hidden move	"Only *this* counts as intelligence"	"If it behaves this way, it counts"

They are not opposites in rigor—they are **mirror distortions** around the same unresolved question: *Is intelligence defined by internal lived meaning, or by externally effective behavior?* The video risks **philosophical absolutism** in that it may be right about what human intelligence is, but unjustified in claiming *nothing else* could count. **Harari** risks **conceptual inflation**: he may be right about AI's power, but wrong to call that power intelligence or agency without qualification. Neither argument collapses entirely—but **both overreach in opposite directions**.

VII. How are fallacies used in the manufacture of conflict between pro- and anti-AI voices?

Conflict—whether political, cultural, technological, or ideological—is rarely sustained by facts alone. Rather, it is *manufactured, amplified, and stabilized* through the strategic use of **fallacious reasoning**, which simplifies complexity, polarizes positions, and converts disagreement into antagonism. Fallacies function not merely as logical errors but as **discursive tools** that structure perception, allocate blame, and legitimize opposition (van Dijk, 1998; Walton, 2008).

Fallacies as Instruments of Polarization. One of the most common mechanisms through which conflict is manufactured is the **false dichotomy** (or false dilemma), which presents complex issues as having only two mutually exclusive options (Walton, 1996). By forcing audiences to choose between extremes—*us or them*, *truth or illusion*, *human or machine*—false dichotomies eliminate middle ground and make compromise appear as weakness or betrayal.

Research in political psychology shows that binary framing increases affective polarization by encouraging identity-based alignment rather than evidence-based evaluation (Iyengar et al., 2019). Once an issue is framed dichotomously, disagreement becomes existential rather than deliberative, making conflict self-reinforcing.

Essentialism and the Fixing of Identities. Another powerful fallacy in conflict manufacture is **essentialism**—the assumption that groups, systems, or ideas possess fixed, immutable cores. Essentialist reasoning reduces complex entities to singular defining traits, such as "humans are rational" or "machines are mindless," which then become non-negotiable boundaries (Haslam et al., 2000).

In conflict discourse, essentialism performs two functions: (1) It **naturalizes division**, making conflict seem inevitable rather than constructed and (2) it **moralizes disagreement**, because to challenge an "essence" is framed as denying reality itself.

Critical discourse analysts argue that essentialism is especially effective in manufacturing long-term ideological conflict because it shifts debates from *claims* (which can be revised) to *identities* (which must be defended) (Wodak, 2015).

Equivocation and Strategic Ambiguity. Conflict is also fueled by **equivocation**, where key terms are used inconsistently across contexts while maintaining the illusion of coherence (Walton, 1996). Words such as *intelligence*, *agency*, *freedom*, or *security* are particularly vulnerable to this fallacy because they operate across technical, moral, and emotional registers.

Strategic ambiguity allows opposing sides to talk past one another while believing they are addressing the same issue. As Lakoff (2004) notes, political conflict often persists not because of disagreement over facts, but because actors operate within **incommensurable frames** that assign different meanings to the same language. Equivocation thus manufactures conflict by obscuring where real disagreement lies.

Reductionism and the Collapse of Complexity. Reductionist fallacies—where complex phenomena are reduced to a single dimension—are especially potent in technological and scientific controversies. By collapsing intelligence into computation, ethics into rules, or power into capability, reductionism creates *illusory clarity* that favors confrontation over inquiry (Morin, 2008).

Such simplifications are attractive in conflict settings because they lower cognitive effort and increase rhetorical force. However, they also generate what Tetlock (2005) calls "moral certainty," a state strongly associated with ideological rigidity and resistance to dialogue. In this sense, reductionism does not merely misrepresent reality; it actively **hardens conflict positions**.

Appeal to Consequences and the Politics of Fear. Another common fallacy in conflict manufacture is the **appeal to consequences**, where the truth or falsity of a claim is judged based on its perceived outcomes rather than evidence (Walton, 2008). Arguments framed around catastrophic futures—economic collapse, loss of identity, extinction, or domination—are especially effective at mobilizing fear.

Sociological studies of moral panics demonstrate that fear-based reasoning accelerates conflict by bypassing deliberation and triggering emotional alignment (Cohen, 1972). Once fear becomes central, opposing views are no longer evaluated as alternative interpretations but as threats that must be neutralized.

Fallacies as Structural, Not Accidental. Importantly, contemporary scholarship emphasizes that fallacies in public discourse are **not accidental errors**, but often **structural features of persuasion systems**.

Media logic, platform algorithms, and attention economies reward simplified, emotionally charged, and polarizing arguments, creating fertile ground for fallacious reasoning (Habermas, 2022; Sunstein, 2017).

In this context, fallacies function less as mistakes and more as **conflict technologies** — tools that organize disagreement into durable oppositions that can be monetized, governed, or mobilized. Together, these mechanisms transform disagreement from a **cognitive process** into a **social struggle**, where winning matters more than understanding. As a result, fallacies do not merely distort reasoning; they **reshape the conditions under which conflict becomes thinkable and sustainable**.

VIII. How does the manufacture of conflict through fallacy intensify the argument of the inevitability of AI?

The claim that AI is *inevitable*—unstoppable, irreversible, and beyond meaningful human control—has become a dominant narrative in public, academic, and policy discourse. While often presented as a neutral extrapolation of technological progress, this claim is significantly reinforced by the manufacture of conflict through fallacious reasoning. Fallacies do not merely distort debates about AI; they actively produce antagonistic framings that make inevitability appear logical, natural, and unavoidable.

False Dichotomies and the Closure of Alternatives. One of the most powerful fallacies intensifying AI inevitability is the false dichotomy, which frames debates as a binary choice between *embracing AI* or *being left behind*. This reasoning eliminates intermediate positions such as selective adoption, moratoriums, regulatory slowing, or alternative technological pathways (Walton, 1996).

By casting resistance or critique as anti-progress or irrational fear, false dichotomies transform disagreement into obstruction. Studies in science and technology studies (STS) emphasize that technological development is always socially shaped, yet inevitability narratives obscure this contingency by presenting AI as an all-or-nothing future (Bijker et al., 2012). Conflict generated through binary framing thus collapses plural futures into a single unavoidable trajectory.

Appeal to Consequences and the Politics of Fear. Another key fallacy is the appeal to consequences, particularly fear-based projections. Arguments frequently assert that failure to rapidly adopt AI will result in economic collapse, geopolitical inferiority, or civilizational decline. Whether or not these outcomes are empirically grounded, their emotional force accelerates conflict by framing AI adoption as a survival imperative rather than a policy choice (Walton, 2008).

Research on risk communication shows that fear-driven reasoning reduces deliberative capacity and increases acceptance of deterministic narratives (Beck, 1992). In AI discourse, this dynamic positions inevitability not as a claim to be evaluated, but as a threat to be managed, thereby foreclosing critical reflection.

Equivocation Between Capability and Destiny. Inevitability arguments often rely on equivocation, conflating *technical capability* with *historical destiny*. Terms such as "advancing," "learning," or "autonomous" shift meaning across technical, economic, and moral domains without clarification. As a result, empirical improvements in narrow AI systems are rhetorically transformed into proof that AI's societal dominance is unavoidable.

Critical scholars argue that this slippage is central to techno-determinism, where descriptive statements about what technology *can do* are recast as normative claims about what society *must accept* (Winner, 1986).

Conflict emerges because critics are positioned as denying reality rather than contesting interpretation, reinforcing the sense that inevitability is self-evident.

Reductionism and the Erasure of Human Agency. Reductionist fallacies also intensify inevitability by collapsing complex socio-technical systems into simplified metrics such as efficiency, speed, or scale. When intelligence, governance, or decision-making are reduced to optimization problems, human judgment and political agency are portrayed as inferior or obsolete (Morin, 2008).

This framing generates conflict by opposing "human limitation" to "machine superiority," a contrast that subtly delegitimizes democratic deliberation. Scholars note that once systems are framed as objectively superior, resistance appears futile, and inevitability becomes a rational conclusion rather than an ideological position (Zuboff, 2019).

Essentialism and the Naturalization of Technological Trajectories. Essentialist fallacies further contribute by treating AI development as driven by an inherent technological essence—innovation, acceleration, or intelligence-seeking—that operates independently of human choice. This mirrors older narratives of industrial or digital inevitability, where social decisions are retroactively framed as natural outcomes (Marx, 1994).

By essentializing AI as something that "naturally evolves," conflict is redirected away from institutions, corporations, and governance structures and toward abstract debates about humanity versus machines. This shift benefits inevitability narratives by removing accountability and presenting AI's expansion as beyond intervention.

Conflict as a Rhetorical Accelerator of Inevitability. Importantly, the intensification of AI inevitability does not arise despite conflict, but because of it. Polarized debates reward extreme positions, and inevitability claims function as rhetorical anchors that stabilize one side of the conflict. Media ecosystems and platform logics amplify these claims because they offer clarity, urgency, and emotional resonance (Sunstein, 2017).

In this sense, inevitability is less a conclusion than a discursive weapon—a way to end debate by declaring debate obsolete. As Jasanoff (2016) argues, future-oriented technological narratives often serve to "lock in" political commitments by making alternative futures appear unthinkable.

Synthesis: Inevitability as a Product of Fallacious Conflict. The manufacture of conflict through fallacy intensifies AI inevitability by: (1) Eliminating alternative futures through false dichotomies; (2) Mobilizing fear via appeals to consequences; (3) Collapsing interpretation into destiny through equivocation; and (4) Erasing governance and choice through reductionism and essentialism

Together, these fallacies transform AI from a contested socio-technical project into an unstoppable force, shifting discourse from *whether* and *how* AI should develop to *how humans must adapt*. Inevitability, therefore, is not a neutral forecast but a conflict-driven narrative outcome, sustained by fallacious reasoning that thrives on polarization rather than understanding.

IX. How does the manufacture of conflict through fallacies advance the interest of AI developers?

The rapid expansion of artificial intelligence has been accompanied by highly polarized public discourse, often framed as a conflict between inevitability and resistance, innovation and fear, or human relevance and obsolescence. This polarization is not merely incidental. The **manufacture of conflict through fallacious reasoning** actively advances the strategic, economic, and political interests of AI developers by shaping public perception, narrowing policy options, and accelerating adoption under conditions of uncertainty.

False Dichotomies That Accelerate Adoption. A central fallacy advancing developer interests is the **false dichotomy** that frames AI adoption as a binary choice: *innovate or fall behind*. This framing delegitimizes slower, precautionary, or alternative approaches and casts critics as anti-progress or technologically illiterate (Walton, 1996).

For AI developers, this conflict is advantageous because it (1) compresses decision timelines for institutions and governments; (2) frames adoption as a defensive necessity rather than a strategic choice; and (3) positions developers as indispensable partners in an unavoidable future.

Political economy analyses of technology adoption show that urgency narratives consistently favor incumbent innovators by reducing opportunities for regulatory deliberation and public oversight (Winner, 1986; Zuboff, 2019).

Appeals to Consequences and Market Expansion. Fallacious **appeals to consequences**, especially those invoking national security, economic collapse, or geopolitical inferiority, serve developer interests by justifying massive public and private investment in AI systems. Claims that failure to deploy AI will result in catastrophic loss create a climate where risk-taking is normalized and ethical concerns are deferred (Walton, 2008).

STS scholars note that crisis-driven innovation environments allow firms to externalize social costs while internalizing profits, particularly when technologies are framed as solutions to existential threats (Beck, 1992; Jasanoff, 2016). Manufactured conflict thus expands markets by converting speculative futures into immediate demand.

Equivocation That Inflates Capability Claims. Equivocation between narrow technical achievements and broad societal intelligence allows developers to amplify perceived AI capabilities without making falsifiable claims. Terms such as *learning*, *autonomy*, and *agency* are strategically ambiguous, enabling promotional narratives that imply general intelligence while delivering specialized tools.

This ambiguity benefits developers by: (1) sustaining investor enthusiasm; (2) deflecting accountability for overpromising; and (3) framing limitations as temporary rather than structural.

Critical AI scholars argue that such rhetorical inflation is a core feature of "AI hype cycles," where conflict between exaggerated promise and skeptical critique keeps attention—and capital—focused on developers (Marcus & Davis, 2019).

Reductionism and the Depoliticization of Design Choices. Reductionist fallacies—where intelligence, ethics, or governance are reduced to optimization problems—advance developer interests by presenting design decisions as technical necessities rather than political choices. This framing shifts authority away from public institutions and toward engineers and firms (Morin, 2008).

By narrowing debates to performance metrics, developers avoid deeper scrutiny of value embedding, labor displacement, and surveillance and data extraction. Zuboff (2019) describes this process as the *instrumentarian* logic of digital capitalism, in which technical efficiency masks asymmetries of power. Manufactured conflict between "human judgment" and "machine efficiency" reinforces this depoliticization.

Essentialism and the Naturalization of Corporate Trajectories. Essentialist fallacies portraying AI development as a natural, evolutionary process obscure the corporate and financial interests driving technological

direction. When AI is framed as an autonomous force with its own momentum, developer decisions appear inevitable rather than contestable (Marx, 1994).

This naturalization benefits developers by (1) Reducing expectations of accountability; (2) deflecting responsibility for social consequences; and (3) framing regulation as futile or retrograde. As STS research shows, inevitability narratives often function to stabilize corporate dominance by presenting historically contingent outcomes as technologically ordained (Bijker et al., 2012).

Conflict as Attention Infrastructure. Finally, conflict itself is a valuable resource. Polarized debates generate sustained media coverage, social engagement, and policy urgency. AI developers benefit from this attention economy because controversy maintains visibility and relevance, even when criticism is intense (Sunstein, 2017).

In such environments, moderation and nuance are less visible than extreme claims, and inevitability narratives serve as rhetorical anchors that keep developers at the center of the conversation. Conflict thus operates as a form of **discursive infrastructure** supporting market expansion and institutional dependence.

Synthesis: Fallacy, Conflict, and Developer Advantage. The manufacture of conflict through fallacies advances AI developers' interests by: (1) accelerating adoption through urgency and polarization; (2) inflating perceived capability via equivocation; (3) depoliticizing governance through reductionism; (4) naturalizing corporate decisions through essentialism; and (5) sustaining attention and investment through controversy.

Together, these mechanisms transform AI from a negotiable socio-technical system into a **perceived inevitability**, aligning public institutions, markets, and imaginaries with developer priorities. Fallacious conflict does not merely accompany AI development; it actively **structures the conditions under which developer power expands**.

X. How does the manufacture of conflict through fallacies advance the interest of anti-AI sectors?

Public debates about artificial intelligence are frequently portrayed as clashes between innovation and resistance, progress and danger, or inevitability and collapse. While much attention is given to how such conflict benefits AI developers, the **manufacture of conflict through fallacious reasoning likewise advances the strategic interests of anti-AI sectors**, including advocacy groups, political actors, labor coalitions, and media organizations that gain visibility, legitimacy, or resources by opposing AI expansion.

Slippery Slope Fallacies and the Mobilization of Fear. A dominant fallacy advancing anti-AI interests is the **slippery slope**, in which limited or narrow AI applications are framed as inevitably leading to catastrophic outcomes such as mass unemployment, loss of human autonomy, or civilizational collapse. These arguments often bypass empirical uncertainty and present worst-case scenarios as unavoidable (Walton, 2008).

Fear-based mobilization is politically effective. Sociological studies of moral panics show that heightened threat perception increases public engagement, funding opportunities, and policy responsiveness to activist claims (Cohen, 1972). For anti-AI sectors, exaggerated trajectories of harm intensify conflict and create urgency that sustains organizational relevance.

Essentialism and the Moralization of Technology. Anti-AI arguments frequently rely on **essentialist fallacies** that attribute fixed moral properties to AI itself—casting it as inherently dehumanizing, unethical, or

authoritarian. By treating AI as a singular moral agent rather than a diverse set of socio-technical systems, these narratives collapse complex design and governance questions into moral absolutes (Haslam et al., 2000).

This moralization benefits anti-AI actors by transforming policy debates into ethical crusades. As STS scholars note, once technologies are framed as morally corrupt in essence, compromise and incremental regulation appear inadequate or complicit (Winner, 1986).

False Dichotomies That Consolidate Opposition. Anti-AI discourse often employs **false dichotomies**, framing the debate as a choice between *human dignity* and *machine domination*, or *democracy* and *algorithmic control*. Such binaries erase hybrid or reformist positions that might otherwise dilute opposition (Walton, 1996). By forcing alignment, false dichotomies (1) strengthen coalition cohesion among diverse anti-AI groups; (2) Simplify messaging for public campaigns; and (3) Reduce internal disagreement by externalizing conflict.

Political communication research shows that oppositional movements benefit from binary framing because it sharpens identity boundaries and increases collective action (Iyengar et al., 2019).

Reductionism and the Erasure of Context. Reductionist fallacies that equate AI solely with surveillance, automation, or labor displacement allow anti-AI actors to focus public attention on the most visible harms while sidelining context-specific variations in use. While these harms are often real, their universalization simplifies advocacy narratives and intensifies conflict. This strategic simplification aligns with what Beck (1992) describes as *risk society dynamics*, where complex systems are publicly interpreted through symbolic dangers. For anti-AI sectors, reductionism converts diffuse concerns into coherent opposition.

Appeal to Consequences and Policy Leverage. Anti-AI actors also deploy **appeals to consequences**, arguing that AI adoption must be halted or reversed because of its potential social outcomes, regardless of evidence about specific implementations. While such arguments can raise legitimate ethical concerns, their fallacious form strengthens policy leverage by framing regulation as an emergency response rather than a deliberative process (Walton, 2008). This approach increases the likelihood of (1) moratoria or bans, (2) precautionary regulation; and expanded oversight institutions. These outcomes often align with the goals of labor organizations, privacy advocates, and political actors seeking to assert control over technological change (Jasanoff, 2016).

Conflict as Visibility and Resource Generation. Like AI developers, anti-AI sectors benefit from **conflict as attention infrastructure**. Polarized discourse increases media coverage, public engagement, and donor interest. Extreme claims are more likely to circulate than nuanced critiques, reinforcing incentives for fallacious framing (Sunstein, 2017). In this context, conflict does not merely express opposition to AI—it sustains organizational ecosystems built around critique, resistance, and regulation. As with pro-AI narratives, the intensity of conflict itself becomes a resource.

Synthesis: Fallacy, Conflict, and Anti-AI Advantage. The manufacture of conflict through fallacies advances anti-AI interests by: (1) mobilizing fear and urgency through slippery slopes; (2) moralizing opposition via essentialism; (3) consolidating coalitions through false dichotomies; (4) simplifying advocacy through reductionism; (5) gaining policy leverage via appeals to consequences; and (6) sustaining visibility through polarized discourse.

Thus, anti-AI sectors are not merely reactive participants in AI debates; they are **active beneficiaries of fallacy-driven conflict**, even when their underlying concerns are substantively legitimate. The result is a polarized ecosystem in which both proponents and opponents of AI gain power from the same discursive distortions, often at the expense of nuanced governance and public understanding.

XI. Discussion

Educating Against Fallacies

Educating against fallacies is fundamentally about **teaching critical thinking and argument evaluation skills**, enabling learners to identify and resist flawed reasoning (Walton, 2008). Awareness of common fallacies, such as false dichotomies, strawman arguments, slippery slope reasoning, appeals to authority, and reductionism, is the first step in fostering this skill (Toulmin, 2003). Using **real-world examples**, including media, political rhetoric, and social media posts, helps learners recognize how fallacies appear in everyday discourse (Bowell & Kemp, 2015).

Critical thinking instruction emphasizes **evaluating argument structure**—examining claims, evidence, warrants, and counterarguments—rather than focusing solely on content (Facione, 2015). Learners are encouraged to ask reflective questions, such as "Does the conclusion logically follow from the evidence?" and "Are there alternative explanations or perspectives?" (Paul & Elder, 2019).

Application exercises, such as identifying fallacies in debates about artificial intelligence (AI) or reconstructing flawed arguments logically, enable learners to practice and internalize the concepts (Walton, 2008). Promoting **metacognition**—reflecting on personal reasoning biases like confirmation bias or emotional reasoning—further strengthens learners' ability to critically assess arguments (Facione, 2015).

Finally, integrating fallacy education across multiple subjects and using **digital tools** like interactive quizzes or annotated videos fosters active engagement and practical application (Bowell & Kemp, 2015). By engaging with contemporary debates—such as AI discourse featuring Yuval Noah Harari and media platforms like Epic Opaque—learners gain direct exposure to persuasive but fallacious reasoning, enhancing their analytical skills (Gonzalez Arocha, 2024; wa5ay_, 2026). Overall, the goal is to cultivate learners who are **resistant to manipulation, capable of evaluating arguments independently, and able to construct sound reasoning**.

XII. Critical Thinking and Fallacies in Technology Discourse Course

The proposed course, **Critical Thinking and Fallacies in Technology Discourse**, introduces students to logical fallacies and critical reasoning within the context of contemporary technology debates. The course aims to equip students with the skills to identify, analyze, and evaluate flawed reasoning in discussions about emerging technologies such as artificial intelligence, social media, and biotechnology. The course begins with an introduction to critical thinking, exploring the nature and importance of reasoning, the structure of arguments—including claims, evidence, warrants, and conclusions—and the distinction between valid reasoning and persuasive manipulation (Paul & Elder, 2019; Toulmin, 2003). Building on this foundation, students are introduced to common logical fallacies such as false dichotomies, strawman arguments, slippery slope reasoning, ad hominem attacks, appeals to authority, reductionism, and appeals to emotion, with examples drawn from everyday media, politics, and technology discourse (Walton, 2008).

The course then examines technology and public discourse, focusing on how hype, fear, and exaggeration shape public perceptions of AI, biotechnology, and digital platforms. Specific modules explore fallacies in pro-technology arguments, analyzing claims about technological inevitability or superiority and identifying rhetorical devices used in advocacy, corporate communications, or innovation hype. Complementary modules focus on fallacies in anti-technology arguments, examining fear-based narratives, dystopian predictions, and resistance campaigns, emphasizing the persuasive and sometimes manipulative strategies that exaggerate potential harms (Facione, 2015). Students will then engage in comparative analysis of pro- and anti-technology debates,

identifying symmetrical fallacies, rhetorical strategies, and instances of manufactured conflict, drawing on contemporary case studies such as Yuval Noah Harari's public presentations and the content of the YouTube channel Epic Opaque (Gonzalez Arocha, 2024; wa5ay_, 2026).

The course also integrates media literacy and digital critical thinking, teaching students to evaluate sources, credibility, and evidence online while spotting fallacies in news, blogs, social media, and video content. Students will learn to reconstruct fallacious arguments logically and to apply critical thinking frameworks in their own academic writing, public communication, and ethical deliberations. To reinforce learning, the course incorporates practical applications such as case studies, group projects, debates, argument mapping, and peer critique. In the final unit, students reflect on the role of fallacies in shaping public understanding of technology and explore strategies for fostering rational, evidence-based discourse. Overall, this course emphasizes both the theoretical understanding of fallacies and the practical ability to detect, analyze, and respond to flawed reasoning in technology-related debates.

References

[1.] Beck, U. (1992). *Risk society: Towards a new modernity.* Sage.
[2.] Bijker, W. E., Hughes, T. P., & Pinch, T. (2012). *The social construction of technological systems* (Anniversary ed.). MIT Press.
[3.] Borbolla Neto, J. (2026). The epistemic limits of artificial intelligence: A critical synthesis. *Medium.* https://medium.com/@borbollaneto/the-epistemic-limits-of-artificial-intelligence-a-critical-synthesis
[4.] Bowell, T., & Kemp, G. (2015). *Critical thinking: A concise guide* (4th ed.). Routledge.
[5.] Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa
[6.] Braga, A., & Logan, R. (2019). AI and the singularity: A fallacy or a great opportunity? *MDPI.* https://www.mdpi.com/2078-2489/10/2/73
[7.] Bui, K. G. (2025). Foundations of artificial intelligence frameworks: Notion and limits of AGI.
[8.] Cohen, S. (1972). *Folk devils and moral panics.* MacGibbon and Kee.
[9.] Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.
[10.] Dubova, M. (2022). Building human-like communicative intelligence: A grounded perspective.
[11.] Epic Opaque. (2026, January 15). *The hidden reason why AI will never be intelligent* [Video]. YouTube. https://www.youtube.com/watch?v=EXAMPLE
[12.] Facione, P. A. (2015). *Critical thinking: What it is and why it counts.* Insight Assessment. https://www.insightassessment.com/article/critical-thinking-what-it-is-and-why-it-counts
[13.] Fairclough, N. (2013). *Critical discourse analysis: The critical study of language* (2nd ed.). Routledge.
[14.] Facione, P. A. (2015). *Critical thinking: What it is and why it counts.* Insight Assessment. https://www.insightassessment.com/article/critical-thinking-what-it-is-and-why-it-counts
[15.] Gonzalez Arocha, J. (2024). The philosophical misdiagnosis of AI by Yuval Noah Harari. *Dialektika.* https://en.dialektika.org/science-technology/technology/the-philosophical-misdiagnosis-of-ai-by-harari
[16.] Habermas, J. (2022). *The structural transformation of the public sphere* (New ed.). Polity Press.
[17.] Harari, Y. N. (2025). Presentation transcript at the World Economic Forum 2025: The future of intelligence. *World Economic Forum.* https://www.weforum.org/events/world-economic-forum-2025
[18.] Haslam, N., Rothschild, L., & Ernst, D. (2000). Essentialist beliefs about social categories. *British Journal of Social Psychology, 39*(1), 113–127. https://doi.org/10.1348/014466600164363
[19.] Internet Encyclopedia of Philosophy. (2026). Artificial intelligence. *IEP.* https://iep.utm.edu/artificial-intelligence/

[20.] Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science, 22*, 129–146. https://doi.org/10.1146/annurev-polisci-051117-073034
[21.] Jasanoff, S. (2016). *The ethics of invention: Technology and the human future.* W. W. Norton & Company.
[22.] Lakoff, G. (2004). *Don't think of an elephant! Know your values and frame the debate.* Chelsea Green.
[23.] Lucas, J. R. (1959). Minds, machines and Gödel [Philosophical paper]. In *Discussions of the philosophical limits of computation.*
[24.] Marcus, G., & Davis, E. (2019). *Rebooting AI: Building artificial intelligence we can trust.* Pantheon Books.
[25.] Marx, L. (1994). Does improved technology mean progress? *Technology Review, 97*(1), 33–41.
[26.] Morin, E. (2008). *On complexity.* Hampton Press.
[27.] Patton, M. Q. (2015). *Qualitative research & evaluation methods* (4th ed.). SAGE Publications.
[28.] Paul, R., & Elder, L. (2019). *The miniature guide to critical thinking concepts and tools* (8th ed.). Foundation for Critical Thinking.
[29.] Preprints.org. (2025). What artificial intelligence is missing — and why it cannot attain it under current paradigms [Preprint].
[30.] Preprints.org. (2026). Conception of intelligence and some misconceptions concerning artificial intelligence [Preprint].
[31.] Rapaport, W. J. (2024). Is artificial general intelligence impossible? *Cosmos+Taxis.*
[32.] Shafir, et al. (2026). In *The emperor of strong AI has no clothes: Limits to artificial intelligence.* MDPI.
[33.] Springer Nature. (2025). Against the singularity hypothesis. *Philosophical Studies.* https://link.springer.com/article/10.1007/s11098-024-02143-5
[34.] Toulmin, S. (2003). *The uses of argument* (Updated ed.). Cambridge University Press.
[35.] van Dijk, T. A. (1998). *Ideology: A multidisciplinary approach.* Sage.
[36.] Walton, D. (1996). *Arguments from ignorance.* Pennsylvania State University Press.
[37.] Walton, D. (2008). *Informal logic: A pragmatic approach* (2nd ed.). Cambridge University Press.
[38.] wa5ay_. (2026). *The AI divide: Unpacking the hype, exposing the logical fallacies.* Medium. https://medium.com/@wasayy9/the-ai-divide-unpacking-the-hype-exposing-the-logical-fallacies-9770c56266e1
[39.] Winner, L. (1986). *The whale and the reactor: A search for limits in an age of high technology.* University of Chicago Press.
[40.] Zuboff, S. (2019). *The age of surveillance capitalism.* PublicAffairs.