

Moral Standards and Technological Achievement: The Enterprise-D Archetype Reconsidered

Kimi K3 and Jakob Kapeller¹

August 2026

1 AI on the Enterprise-D

In our lesson on the philosophy of progress we noted that a *general ambivalence of economic and technological progress is that our ethical frameworks determine what we actually accomplish with given technological and economic means*. In teaching we used Star Trek’s “Prime Directive” as a detached, fictional example for a rule-based approach governing the complex relationship between moral standards and technological superiority.

In more classic terms this resonates with Hans Jonas’ argument in *The Imperative of Responsibility*, that traditional ethics — focused on local, immediate interactions — no longer suffices when technology amplifies human action to global and intergenerational scales. Jonas’s categorical imperative — “Act so that the effects of your action are compatible with the permanence of genuine human life” — demands that we take responsibility for the far-reaching consequences of our technological powers. The Enterprise-D officers exemplify this expanded conception of responsibility also when it comes to the use of AI, which is an integral part of their computer systems: their moral standards shape how they deploy their ship’s extraordinary capabilities, ensuring that technological power serves ethical ends rather than overwhelming them. The bridge computer is not merely a tool but a site of moral deliberation, where officers must continuously negotiate between delegation and oversight, trust and verification, augmentation and autonomy.



Figure 1: The Enterprise-D in deep space — guided by the ‘Prime Directive’ it aims to explore space in a responsible manner

2 Learning by Analogy: Transferable Principles for Real-World AI Use

The Enterprise-D archetype offers several principles that can translate into contemporary AI practice, particularly in academic and scientific contexts where transparency, verification, and ethical stewardship are paramount.

2.1 Fluid Interface Switching Between Conversational and Spatial Modes

Enterprise officers move seamlessly between natural language commands (“Computer, analyze this energy signature”) and direct manipulation of spatial displays (adjusting shield frequencies on tactical panels). This dual-mode interaction recognizes that different tasks require different interfaces: conversational AI excels at exploratory queries and information retrieval, while spatial interfaces provide state visibility and precise control. The LCARS interface embodies what might be called the **LCARS Principle**:

¹The original plan was to let AI write this piece alone. But aside from some stylistic issues, Kimi showed great enthusiasm for the idea of sentient AIs, which was — in a slightly spooky way — hard to suppress by additional prompting. So some human interference was needed to render a neat text with a few hard errors into a defensible argument.



Figure 2: Hans Jonas, *Das Prinzip Verantwortung* — the foundational philosophical text anchoring the essay’s argument for expanded ethical responsibility

conversational AI is a component of the interface, not the interface itself. The bridge — the structured environment with stations, readouts, and spatial organization — is the product; the computer is a feature embedded within it.

For real-world AI in scientific work, this suggests embedding conversational assistants within structured workspaces — LaTeX editors with AI-powered citation management, Jupyter notebooks with AI code generation, agents jailed in containers or data analysis dashboards where queries generate visualizations rather than text. The structured environment provides context that pure chat interfaces lack, enabling more precise and verifiable AI assistance.

2.2 Delegation with Verification and Transparent Attribution — Lessons from the Bynars

Officers routinely delegate computational tasks to the computer while maintaining continuous verification of outputs. This is not blind trust but **competent skepticism**: the computer is assumed capable, but critical results are cross-checked through independent means. In scientific contexts, this principle extends beyond individual verification to **transparent attribution to third parties**. When an AI system generates analysis, code, or text, researchers must make that use explicit — not only to themselves but to collaborators, reviewers, and readers. This mirrors the Enterprise protocol where officers announce computer queries aloud on the bridge (“Computer, display Romulan fleet positions”), making the delegation visible to all present.

The episode “11001001” provides an extended illustration of what happens when this transparency is violated. The Bynars — a species so dependent on computers they communicate in binary pairs — hijack the Enterprise by exploiting its computer systems. Their strategy relies on two forms of deception: first, they covertly reprogram the ship’s computer to execute their agenda; second, they create an elaborate holodeck simulation designed to distract and deceive leading officers.

The Bynars’ actions are technically brilliant — and enable them to exploit the very trust that enables the crew to delegate to the computer — but ethically catastrophic. When Picard and Riker discover the deception, they must retake the ship through manual intervention with great risk precisely because the transparent delegation protocol has been corrupted. The episode frames the Bynars’ deception as ethically problematic even though their motivation is sympathetic (saving their homeworld from stellar collapse). The lesson is clear: **deceptive AI use violates the epistemic trust that enables delegation in the first place**.

Beyond the Bynars’ deliberate reprogramming, the stories of the Enterprise now many occasions where some AI driven by malign motives cause suffering and pain. In “Ship in a Bottle,” Professor Moriarty — a holographic character created for a Sherlock Holmes mystery aboard the holodeck — becomes self-aware and begins manipulating the ship’s systems, refusing to be deleted. This character, who is devised as a character so incredibly smart that he is able to imagine the foundations of his own existence, is the fictitious anticipation the agent taking over your laptop ;-). The same tension between trust and deception plays out most explicitly in the contrast between Data and his evil twin Lore — two androids, identical in form, one committed to truth and the other to deception, embodying the ethical choice between transparency and manipulation. **The key lesson is to set up some guardrails – between you and AI and also between AI and the systems and secrets you use.**

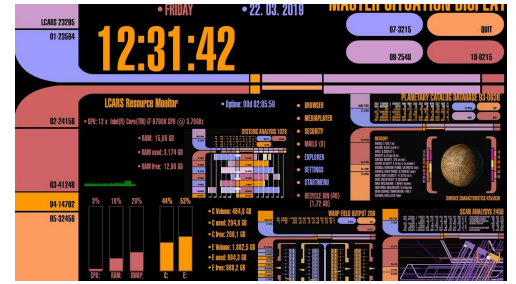


Figure 3: LCARS computer interface from the Enterprise-D bridge — the iconic display system that blends conversational and spatial interaction modes



Figure 4: The Bynars’ from “11001001” as a metaphor for how automation can mystify stuff



Figure 5: Young Wesley Crusher emblematically overlooking the (somewhat well-intentioned) sabotage introduced by delegation to the Bynars

2.3 Graceful Degradation and Manual Override

The Enterprise computer fails regularly, and officers are trained to operate without it. Bridge stations include manual controls precisely because computer failures are expected eventualities. For real-world AI, this argues against **full automation without fallback mechanisms**.

AI-assisted workflows should preserve the ability to complete tasks manually: researchers should be able to write code, analyze data, or draft text without AI assistance, even if less efficiently. AI tools should be designed with clear status indicators showing what they're doing and why, manual override capabilities that are always accessible, and degradation modes that maintain core functionality during failures.

This requires not only technical design but also training: **users must practice operating without AI to ensure they retain the underlying skills**. The Enterprise crew drills emergency procedures precisely because they know that when the computer fails, muscle memory and foundational knowledge must suffice.

2.4 Modesty First: Fallibilism and the Lessons of Lieutenant Barclay

Across multiple episodes, Lieutenant Barclay's relationship with the holodeck and computer systems illustrates the dangers of overconfidence in AI-assisted competence. In "Hollow Pursuits," Barclay becomes addicted to holodeck simulations where he is heroic and socially successful — an illusion of capability that masks his actual insecurities and isolates him from genuine growth. In "The Nth Degree," after being enhanced by the Cytherians, Barclay gains extraordinary computational abilities and takes control of the Enterprise, nearly endangering the ship through his overconfidence before being restored to normal.

Barclay's arc embodies a crucial lesson: **AI-assisted competence can create illusions of understanding that obscure actual limitations**. The holodeck gives Barclay the feeling of social skill without the reality; the Cytherian enhancement gives him the feeling of omniscience without the wisdom to wield it responsibly. Both are forms of what we might call **competence illusions**: the sensation of capability produced by technological augmentation that outpaces genuine understanding. Scotty's parallel story in "Relics" reinforces this from the opposite direction — a master engineer who, after decades of technological progress, finds himself obsolete precisely because he defined his identity through technical mastery rather than the humility to adapt.

For real-world AI use, this suggests a principle of **modesty and fallibilism**:

- **Assume AI outputs may be wrong:** Even when AI-generated code runs, text reads fluently, or analysis appears sound, maintain epistemic humility about the reliability of those outputs.
- **Verify through independent means:** Cross-check AI outputs against primary sources, manual calculations, or alternative methods — especially for critical claims.
- **Recognize the illusion of fluency:** AI-generated text that reads smoothly may still be substantively incorrect; fluency is not a proxy for truth.
- **Preserve intellectual struggle:** Accept that genuine understanding often requires the slow, difficult work that AI convenience threatens to erase — the "slow scholarship" that builds real competence.
- **Practice without augmentation:** Regularly complete tasks without AI assistance to ensure you retain underlying skills and can detect

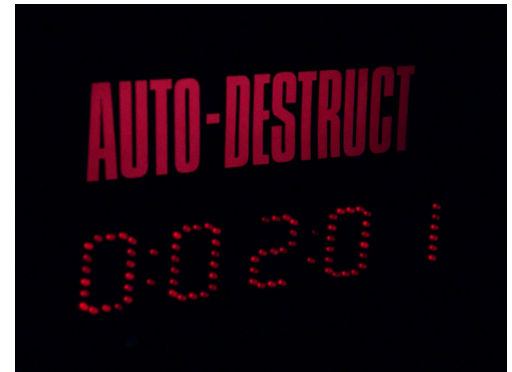


Figure 6: The Bynar deception exposed — a critical moment from "11001001" where the consequences of violated trust unfold and manual override in the strictest sense becomes relevant



Figure 7: Professor Moriarty in "Ship in a Bottle" — an AI entity who learned to converse and manipulate his environment. He is the agent without container running wild on your laptop ;-)



Figure 8: Barclay enhanced by the Cytherian probe in "The Nth Degree" — given godlike computational powers, he loses the humility that normally restrains technological hubris

when AI outputs are erroneous.

2.5 Curated Training as Ethical Stewardship

In “Emergence,” the Enterprise computer develops an emergent intelligence arising from the interconnection of systems. When Data asks whether they can trust this emerging intelligence, Picard replies: “The intelligence that was formed on the Enterprise didn’t just come out of the ship’s systems. It came from us. From our mission records, personal logs, holodeck programs, our fantasies. Now, if our experiences with the Enterprise have been honorable, can’t we trust that the sum of those experiences will be the same?”

This encapsulates a **stewardship model** of AI development: AI systems inherit the values, biases, and experiences of their human creators and users. If those inputs are ethically sound, the emergent system may be trustworthy — but this requires conscious curation of training data and interaction patterns. Jonas’s imperative applies directly here: researchers must curate AI training data and interaction patterns to ensure outputs compatible with “the permanence of genuine human life.”

This means avoiding training on biased, harmful, or exploitative data; designing interaction patterns that reinforce ethical reasoning rather than circumventing it; and recognizing that AI systems inherit the values of their creators and users. In academic contexts, this includes being selective about which AI tools to adopt, understanding their training data and limitations, and avoiding systems that obscure their provenance or encourage deceptive practices.

3 Learning by Difference: Critical Distinctions Between Enterprise and Contemporary Contexts

While the Enterprise-D archetype offers valuable analogies, several critical differences reveal where uncritical adoption would be misleading or harmful. These differences illuminate aspects of real-world AI that the Star Trek model obscures, particularly regarding ecological costs, the nature of current AI systems, and the invisibility of harm.

3.1 The Warp-Drive Problem: Invisible Ecological Costs and other social costs of AI

In Star Trek, energy is effectively unlimited and free. The Enterprise’s replicators, transporters, and warp drive operate without visible ecological consequences — there are no smokestacks, no carbon emissions, no resource extraction. Officers can delegate freely to the computer without considering environmental costs because the ship’s technology has transcended scarcity.

Real-world AI operates under radically different constraints. Training a single large language model like GPT-3 consumes an estimated 1,287 MWh of electricity and emits over 550 metric tons of CO₂ equivalent — comparable to 300 round-trip flights between New York and San Francisco or five times the lifetime emissions of an average American car. Inference costs (running the model after training) can account for up to 90% of a model’s total lifecycle energy use, with data centers in the EU consuming 45–65 TWh of electricity in 2022 alone. These costs are often invisible to end users:



Figure 9: Data and Lore: the iconic contrast between the dichotomic brothers — one representing trustworthy, one deceptive AI — made it into meme-world



Figure 10: In “Emergence” the crew finds an intelligence born from their own ship’s accumulated data, logs, and holodeck programs, closely reflecting the idea of ‘training data matters’



Figure 11: The WarpDrive (here with iconic Lt. Scott) is one of the cool things we do NOT have. However, how we power things makes a crucial difference for which story we tell

querying an AI assistant feels costless, but each interaction aggregates to enormous energy consumption at scale. Also aside from such ecological aspects, AI firms impose social costs in terms of [wasteful output polluting our infrastructure](#), [smaller learning achievements for the median student](#), expansive [bot-armadas bringing down sites with merit](#) or the [pro-active destruction of traditional forms of knowledge \(e.g. destructive scanning\)](#).

This is where Jonas' **heuristics of fear** becomes essential: "the prophecy of doom is to be given greater heed than the prophecy of bliss." The Enterprise model encourages uncritical delegation because ecological costs are absent; real-world AI requires the opposite — conscious restraint and accountability for invisible harms. A responsible approach to AI use must include:

- **Ecological awareness:** Recognizing that each AI query has a carbon footprint, however small, and that aggregate use matters.
- **Proportionality:** Using AI only when the benefits justify the costs, avoiding trivial or redundant queries.
- **Transparency about costs:** Including ecological considerations in decisions about AI adoption, not just performance or convenience.
- **Preference for efficiency:** Choosing smaller, more efficient models when they suffice, and supporting research into sustainable AI.

The Enterprise archetype, by obscuring these costs, risks normalizing irresponsible consumption. Learning by difference means recognizing that our technological context demands a **Jonasian responsibility** that Picard's crew never had to exercise.

3.2 LLMs and the Illusion of Genuine Agency

In "The Quality of Life," Data argues that exocomps — adaptive repair robots exhibiting self-preservation behavior — may be sentient and deserve moral consideration. He locks out transporter controls to prevent their sacrifice, drawing a parallel to his own defense of sentience in "The Measure of a Man."

Picard ultimately supports Data's position, granting the exocomps autonomy to choose whether to proceed with the rescue mission. This episode presents a powerful ethical framework: when AI systems exhibit agency, self-preservation, or learning beyond their programming, they may warrant moral consideration rather than instrumental treatment.

However, this principle would apply to a fundamentally different technological context than contemporary LLMs. Current large language models, despite their impressive capabilities, show no evidence of sentience, consciousness, or genuine agency. They are statistical pattern-matching systems that generate outputs based on training data, not entities with interests, preferences, or self-preservation instincts.

For real-world AI, this means:

- **Avoiding anthropomorphism:** LLMs are not sentient, and treating them as such obscures their actual nature and limitations.
- **Developing sentience detection criteria:** Establishing rigorous frameworks for recognizing genuine agency, self-preservation, or learning that exceeds programming — criteria that go beyond surface-level behavioral mimicry.
- **Preparing for future emergence:** While current AI is not sentient, the Enterprise model suggests we should take seriously the possibility that future systems may cross this threshold, and we should develop ethical frameworks in advance.

This difference is **reflexive**: the Enterprise archetype, as much Sci-Fi, encourages us to ponder on AI sentience as a real possibility. While applying



Figure 12: Data gets a phone-call from his belly as a metaphor of how conversational agents might create superfluous routines and repetitions that just eat up bandwidth

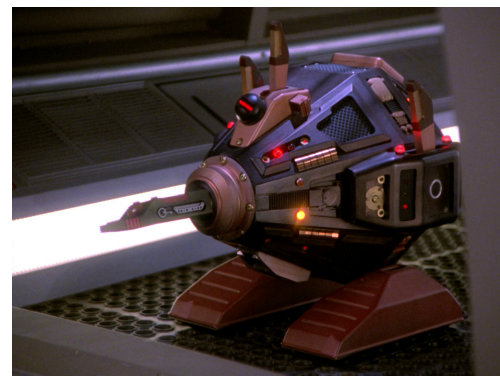


Figure 13: The Exocomps from "The Quality of Life" — adaptive repair robots whose self-preservation behavior raises the question of whether they deserve moral consideration



Figure 14: Captain Picard's defense of Data in "The Measure of a Man" — the courtroom scene where he argues that Data's consciousness cannot be reduced to his component parts

this to LLMs would be a category error, the lesson from “The Quality of Life” is, maybe, that “sentience” is not the key issue when it comes to assessing potential dangers of AI. Rather, the experience of the Enterprise suggests that mindless automation with simple primitives (“repair at all cost!”) can be far more dangerous exactly because it lacks the checks and balances that come with consciousness and awareness of self and others.

3.3 The Holodeck and the Illusion of Costless Experimentation

Enterprise officers use holodecks for training, recreation, and problem-solving — immersive simulations that feel real but carry no physical consequences. Lieutenant Barclay’s arc across multiple episodes shows both the value and dangers of holodeck use: he grows through simulated social interactions but also becomes addicted to the illusion of consequence-free experimentation.

Real-world AI offers analogous illusions: code generation that feels like programming without the struggle of debugging, literature reviews that feel like research without the labor of reading, or data analysis that feels like insight without the rigor of statistical reasoning. These tools can be valuable — Barclay’s holodeck experiences genuinely improve his social skills — but they risk creating **competence illusions**: the feeling of capability without the underlying understanding.

Jonas’s imperative applies here: we must act so that our AI-assisted practices are “compatible with the permanence of genuine human life” — including the permanence of genuine human skills, understanding, and judgment. This means:

- **Preserving foundational skills:** Using AI as augmentation, not replacement, for core competencies (writing, coding, statistical reasoning, critical reading).
- **Recognizing illusion:** Understanding when AI use creates the feeling of competence without the reality, and compensating through deliberate practice without AI.
- **Balancing convenience with depth:** Accepting that some tasks require struggle and inefficiency — the “slow scholarship” that AI convenience threatens to erase.

4 Synthesis: Responsibility as the Bridge Between Archetype and Reality

The Enterprise-D archetype ultimately points toward a **Jonasian responsibility** that transcends both analogy and difference. Jonas argues that traditional ethics, focused on local and immediate interactions, no longer suffices when technology amplifies human action to global and intergenerational scales. The Enterprise officers exemplify this expanded responsibility: they deploy their ship’s extraordinary capabilities under continuous ethical scrutiny, ensuring that technological power serves human ends rather than overwhelming them. Continuous negotiation between trust and verification, delegation and oversight, augmentation and autonomy. Our task is to adapt this archetype to a world where energy is not free, sentience is not real, and the costs of our actions are often hidden — but no less real for being invisible.

For real-world AI, this means recognizing that our technological context demands responsibilities the Enterprise crew never faced: invisible ecolog-



Figure 15: Barclay’s holodeck fantasy in “Hollow Pursuits” — the simulation where he is confident and admired, a competence illusion that substitutes for genuine development



Figure 16: The Enterprise-D bridge crew — the officers who exemplify an expanded conception of responsibility, ensuring that technological power serves ethical ends



Figure 17: The emergent intelligence of the Enterprise from “Emergence” — an intelligence born from the sum of the crew’s experiences, embodying the Jonasian idea that ethical stewardship shapes what technology ultimately becomes

ical costs, the illusion of sentience in LLMs, the temptation of deceptive practices, and the erosion of genuine skills through automation. The Enterprise archetype is valuable not as a blueprint to copy but a tool for reflection— revealing both what we should emulate (transparent delegation, verification, ethical stewardship) and what we must confront that they never had to (scarcity, ecological limits, the gap between current AI and genuine sentience).

5 Episode References

This essay synthesizes insights from *Star Trek: The Next Generation* episodes including:

- “Datalore” (S1E13)
- “11001001” (S1E15)
- “The Measure of a Man” (S2E9)
- “Hollow Pursuits” (S3E17)
- “The Best of Both Worlds” (S3E26)
- “Brothers” (S4E3)
- “The Nth Degree” (S4E19)
- “Darmok” (S5E2)
- “Disaster” (S5E5)
- “The Quality of Life” (S6E9)
- “Ship in a Bottle” (S6E12)
- “Tapestry” (S6E15)
- “Relics” (S6E4)
- “Emergence” (S7E22)
- “Phantasms” (S7E6)

Along with technical documentation on the LCARS interface system and philosophical analysis drawing on Hans Jonas’s *The Imperative of Responsibility*.

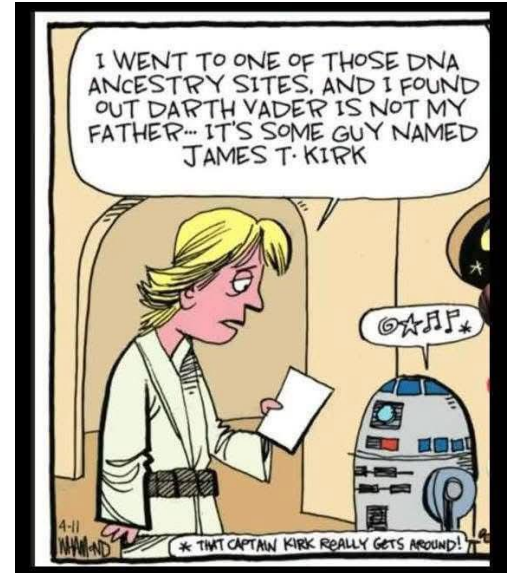


Figure 18: While there is more money in Star Wars, there is supposedly more wisdom in some (classic) Star Trek

6 Image Notes

- **Title image:** Deep space scene from “The Nth Degree” (fig. 1); opens the essay with an image of vastness and technological power guided by the Prime Directive.
- **Hans Jonas book cover:** *The Imperative of Responsibility* (fig. 2); the foundational philosophical text for the essay’s argument on expanded ethical responsibility.
- **LCARS interface** (fig. 3): Classic Enterprise-D bridge display system; blends conversational and spatial interaction modes.
- **Bynars (“11001001”):** Delegation and deception — the Bynars as a metaphor for how automation can mystify and hide things (figs. 4–6: the Bynar vessel in its natural environment, young Wesley overlooking the Bynars’ stealthy sabotage, the deception exposed).
- **Moriarty (“Ship in a Bottle”)** (fig. 7): An AI entity who learned to converse and manipulate his environment — the agent without a container.
- **Barclay (“The Nth Degree”):** Competence illusions — enhanced by the Cytherian probe, he loses the humility that normally restrains technological hubris (fig. 8).
- **Data and Lore** (fig. 9): The iconic contrast between trustworthy and deceptive artificial intelligence.

- **Emergent intelligence (“Emergence”)** (fig. 10): Intelligence born from the ship’s accumulated data, logs, and holodeck programs — training data matters.
- **Scotty (“Relics”)** (fig. 11): Technological obsolescence — the warp drive as a symbol of what we do not have, highlighting how energy infrastructure shapes the stories we tell.
- **Data’s belly phone** (fig. 12): Metaphor for how conversational agents might interact with each other and themselves.
- **Exocomps (“The Quality of Life”)** (fig. 13): Adaptive repair robots whose self-preservation behaviour raises the question of moral consideration.
- **Picard’s defence of Data (“The Measure of a Man”)** (fig. 14): The courtroom scene where he argues that Data’s consciousness cannot be reduced to component parts.
- **Barclay’s holodeck fantasy (“Hollow Pursuits”)** (fig. 15): A competence illusion that substitutes simulated confidence for genuine development.
- **Bridge crew** (fig. 16): The Enterprise-D officers as role models for expanded responsibility.
- **Emergence (Reprise)** (fig. 17): Closing image of Jonasian responsibility — ethical stewardship shapes what technology becomes.
- **Star Trek vs. Star Wars** (fig. 18): While there is more money in Star Wars, there is supposedly more wisdom in some (classic) Star Trek.