Module 9
White Paper: Consensus on Artificial Intelligence

Though no module was dedicated to the topic of artificial intelligence, it was nevertheless a common theme in our discussions and collective inquiries. Extrapolating from the discussions, we present here a dedicated paper to the consideration of AI. In the background of this discussion is the belief that "Artificial Intelligence" is a misnomer predicated on the basis of a fundamental misunderstanding of intelligence itself; this background belief (and the proper understanding of intelligence) can be derived from the materials of Module 2. A more accurate name would be "Intelligence Simulator". Nevertheless, we use the accepted nomenclature of AI for the sake of linguistic continuity—with the admonition that we must strive to rename the technology more accurately in the long-term.

The sections of this paper are more limited than the previous eight—as the discussion was dispersed across all eight modules, we keep our aims here limited to identifying the principal theoretical difficulty posed by AI technologies.

I. Theoretical Considerations

The stance we take on AI becomes clear if we view LLM technology through the two lenses that recur throughout the lectures and discussion: Thomistic psychology and Heidegger's account of technology as a mode of revealing. These lenses do not merely add "ethical concerns" to an otherwise technical subject but attempt to re-situate AI within a more basic question: namely, what does it means to know, to make choices, and to inhabit a world in which we may thrive only by taking responsibility for ourselves and our own actions?

From the Thomistic perspective on the human psyche, human cognition cannot be reduced to a self-contained process of computation that can be replicated by cycles of up-scaling pattern-recognition (the fundamental architecture, that is, of LLMs and so-called "machine-learning"). Human knowing is an act of a living subject ordered to immaterial, universal truth. To understand the deficiency in AI relative to human intelligence, we must understand this claim concerning the ordering of our minds: that the truths we know are immaterial and universal, that is, does not claim a mystical power, but rather recognizes the manner in which we *know*. Crucial here is a certain disambiguation of the words "know" and "knowledge". The English language comprises a multitude of distinct but related operations under this singular word. Other languages, such as Greek, Latin, German, Portuguese, and many others, have employ a multitude of words with distinct nuances of meaning where we, in English, commonly use the word "know": e.g., *scire* and *cognoscere*, *ἐπιστήμη* and *γιγνώσκειν*, *wissen* and *kennen*, *entender* and *conhecer*, among many others. Common to these particular pairings is a distinction between the scientific

knowledge which represents a deep habit of truly understanding our objects and the knowledge of familiarity, by which we have an experience of something without necessarily having attained the specific degree of scientific understanding proper to the heights of intellectual knowing.

It is this scientific sense in which we mean that someone knows immaterial and universal truth. One may *know*, in the sense of familiarity, the feeling of being in a moving car; this is quite different from knowing how a car works. So too, someone may *know* another human being from long years of coexistence; but this is different from *knowing what a human being is*. The latter kind of knowledge consists in something universal and irreducible to the material particular instances. The working of a car (with more variation, given that it is an artifact) consists in certain principles of energy storage and transfer to common mechanisms of motion. The being of a human consists in the ordered disposition of diverse faculties in a kind of body. The same principles govern a Toyota and a Ford—and at a greater degree of abstraction, even a Tesla or Rivian—while the faculties of a human being whether male or female, young or old, Eastern or Western are all fundamentally the same, although subject to diverse instantiation and development.

Put otherwise, we have the capacity to form objects of understanding which are *fundamentally irreducible to particulars*. While we always understand those objects as somehow *through* particulars, we see that the object actually understood is neither any one of them nor the totality or sum of them.

The human intellect, that is, is more than a processor that receives inputs and generates outputs. It is a complex faculty for receiving the intelligible forms of things, discerning those forms as governing of the sensible particulars in which they are discovered, and subsequently judging of the world in light of ultimate ends—it comprises a potential and illuminate infinitude in which meaning is not only found in the objects but coalesces in the life of the self. This intellectual faculty is inseparable from the will, which is ordered to the good, and from the passions and bodily powers that condition our sense-perceptual cognition, our memories, and our moral and psychological habituation. In short, "intelligence" in the human sense includes formal understanding, judgment, and purposive self-directed orientation within an integrated organism. The nature of the human being was discussed in relative brevity, but at far greater length than this, in Module 2 and its associated theoretical paper, "The Soul and the Machine".

AI, by contrast, is being conceived and used as a technological extension of an already pervasive and deleterious modern tendency: to treat cognition as if it were primarily information-processing. LLMs and related systems are powerful because they exploit statistical regularities in human language and behavior, producing plausible continuations and predictions, and simulating natural intelligence in the process. But in Thomistic terms, such systems do not **know** or *have* **knowledge**; they do not possess the intentional "aboutness" of the intellect ordered to truth as truth, to those immaterial and universal objects that render causal explanatory possibilities. They only simulate the outward marks of discourse while lacking the inward act of living understanding that unites concept to reality through judgment. This deceptive imitation matters not merely for accurately identifying the diverse natures, but for the cultural consequences: once a society grows accustomed to treating competence in symbolic

manipulation as intelligence, it becomes easier to treat truth as a mere function of coherence and apparent fluency or popular resonance rather than correspondence to reality.

Emphasis upon well-formed habits therefore re-enters the picture as essential. Thomistic psychology foregrounds the role of *habitus*: stable qualities of the mind that deepen our perceptions, understanding, capacity for judgment, illuminate our desires, and shape the apparent courses of action we may take. Since our technologies do not merely assist in the performance of discrete tasks but mediate in the formation of habits (by both extension and reflexion of our faculties), AI systems—precisely because they operate superficially through language (the medium of common life)—have an unusual power to specifically reshape our intellectual habits. They can tempt the user away from the inherently slow-developing practices of cognitive discipline by which the intellect is formed—close reading, purposive memorization, extended argumentation, prolonged contemplation, rhetorical reflection—and toward a habit instead of continual outsourcing: the assumption that understanding consists in retrieving or generating plausible texts or the right sequence of symbols, rather than in intellectually noetic judgment about *what is* and *what ought to be*.

The seminar therefore treats AI as most dangerous primarily when it becomes accepted as *noetically authoritative*: that is, when it is relied upon not as an instrument subordinate to human judgment or discernment, or used merely as an instrument stimulative for refining our inquiries, but as a surrogate for our intellectual faculties and operations themselves. In Thomistic terms, this is a radical of disordering of our powers. The intellect's acts cannot be delegated without substantive loss to human flourishing, because these acts are not mere input-filtering and output-selection. Rather, they are the living acts by which a rational agent discovers and assents to what is true and rejects what is false, and, moreover, by which the human being builds up the habits of personality. The will, likewise, cannot be reduced to preference satisfaction or probabilistic optimization, because it is ordered to the good as such, and therefore requires persistent and enduring formation—achieved through the identification of ends and development of virtues. AI can assist deliberation as an extrinsic tool of stimulation, but it cannot supply the *measure* whereby deliberation is rightly undertaken nor by which choices are made.

Heidegger provides the seminar with a complementary, and in some ways even more radical, diagnosis of AI's dangers. The core Heideggerian insight is that technology consists not primarily in the collection of tools but as a mode of revealing: a way in which beings come to appear, and in which thinking itself is constrained through a certain disproportioned extension (to supply explicitly, from McLuhan, what Heidegger indicated implicitly). Modern technology, in Heidegger's account, reveals the world as *standing-reserve*—as resources to be ordered, stored, predicted, and made available. This technological form of rationalization does not begin with machines; rather, the machines are invented as instruments rapidly accelerating this inherently inhuman technological mode of disclosure.

In fact, AI presents, when viewed under lights of the conceptual framework provided by Humanitas Technica, a historically unparalleled intensification of this technological revealing. Industrial technology rendered biological and geological nature as energy and resource, but AI renders human life itself—especially in its specifically linguistic and social dimensions—as data to be stored and

manipulated: and from that data, patterns to be modeled, predicted, and optimized. In the digital age, the world appears as a complexifying field of signals and probabilities, and within this field, human beings appear as predictable systems whose behavior can be controlled. Heidegger warned that, in the end, the technological manner of thinking transforms even the human person into a resource. Recognition that AI shifts the grounds towards this fundamentally-inhuman reconception of human behavior and human beings themselves does not mean we are accusing the developers behind these technologies of participating in or deliberately constituting some grand conspiracy. Rather, this recognition identifies the deeper problematic acceptance of the intrinsic governing rationalization belonging to the technological mode of thinking, a mode of thinking that seeks **control** over all things.

In Heidegger's terms, AI in its present forms, usages, and developmental trajectory extends the totalization of **enframing** (*die Gestell*): the demand that reality present itself in a calculable form. Once AI becomes central to institutional decision-making—in education, workforce hiring, medicine, finance, governance, and so on—institutions begin to privilege what is legible to AI systems; this is already happening and presumably will only intensify over the next several years. Even more rapidly than has already been the case for some decades, the criteria of reality will shift toward what can be quantified, categorized, and predicted. That shift then feeds back into human self-understanding: people learn to present themselves in machine-readable ways, optimizing themselves not in a manner proportionate to the human good but rather for system visibility (e.g., SEO and social media algorithm-gaming—and now extend those algorithmic modes of discovery and interpretation into every dimension of human life). Citizenship, workplace behavior, and even the crafting personal identity (e.g., profiles on dating apps) are reorganized by the demand to be legible to the systems and their pattern-recognition algorithms.

This is why the seminar in every module repeatedly distinguishes between harms that are at the "content-level" (such as misinformation, ideological basis, and erroneous interpretation) and harms that are "form-level" (the reshaping of environments through which attention, memory, authority, the conditions of choice, and the habituation of the psyche). The Heideggerian analysis, paired with a Thomistic understanding of the human person, makes precise this form-level understanding of technological harm, especially with reference to AI. In other words, AI does not necessarily deliver bad information or create bad content (it may, in fact, deliver very good and very useful information, and be used to create good content). But it nevertheless, in its present iterations and trajectories, promotes a world in which the primary question becomes, "What can be predicted and controlled?" rather than "What is true?" or "What is good?"

As a consequence, all thinking becomes increasingly channeled into this narrowing pathway, a pathway that leads away from the proper human good. Thus the Thomistic and Heideggerian lenses converge our analysis on a single concern: namely, that AI tempts human beings to accept a substitute for the capacity for intellectual understanding with depth (including memorative depth) and subsequently for making good judgments. For Aquinas, the operation of judgment is a personally-lived act ordered to the truth about the good; for Heidegger, genuine thinking is a receptive openness to being that resists

reduction to calculative manipulation. AI excels at calculative operations in the medium of language and thereby, if we outsource our thinking to it (as we so easily might, given that medium), such acceptance may rapidly atrophy our capacities for discerning the truth. Once this occurs, we thereby allow it (explicitly or implicitly) to determine for us what constitutes "the good". The danger, in other words, is that we come to treat calculative fluency as a sufficient substitute for our own understanding, and predictive power as adequate for human judgment. We would by this process undo our very selves, our very humanity.

From this convergence, the seminar's practical orientation follows naturally: not to declare AI "evil" nor to deny its legitimate uses. But we must refuse the eminently possible inversion by which AI becomes the measure of the human. AI must be kept subordinate to domains that require human judgment: moral deliberation, education understood as formation, political prudence, and any inquiry ordered to truth rather than mere functionality. That subordination is not accomplished by policy or programming alone. It requires virtue, habit, and institutional practices that preserve silence, duration, and responsibility—conditions within which intellect and will can operate integrally.

In short: AI is a particularly potent instance of modern technology's deeper tendency to reshape what we take knowledge and thinking to be. Thomistic psychology clarifies what is lost when understanding is reduced to output and when judgment is outsourced; Heidegger clarifies how an entire world can come to appear primarily as standing-reserve, including human language, communication, experience, and social life. Together, they explain why the question of AI is not merely technical or ethical, but metaphysical and anthropological: it concerns the kind of beings we are becoming as we allow our mode of revealing—and our habits of mind—to be governed by totalizing systems of prediction and control.

Practical Actions Indicated

1. AI must remain subordinate to human judgment, not replace it

In practice, this means AI should be used as an aid to deliberation or analysis, never as the final arbiter in domains requiring prudence—education, hiring, medical decisions, governance, or moral reasoning. Responsibility must remain clearly attributable to persons. Any act of outsourcing properly human judgment to AI constitutes a rejection of humanity and an act hostile to the well-being of the individual, the community to which he or she belongs, and of humankind as a whole.

2. Institutions should preserve non-automated domains of decision-making

The discussions of our seminar consistently stressed that some practices must remain intentionally inefficient: slow reading, deliberative assessment, face-to-face judgment, and discretionary authority. These are not failures of optimization but safeguards for truth, formation, and accountability. It belongs to institutions in possession of authority—and the persons charged with exercising that authority—to maintain environments or domains in which human prudence is cultivated through these practices, and decision-making thereby retain its fundamentally human character.

3. AI use should be limited in formative contexts

Particular caution is required where habits of mind are being formed—especially in education and early professional training. Reliance on AI for thinking, writing, or synthesis risks deforming intellectual habits by substituting output generation for understanding. While someone may use such technologies in a beneficially-stimulative way, such stimulation can be fruitful only to habits already well-formed; otherwise (and even then, if used carelessly), it risks displacing those habits and enervating the human person in the process.

4. Surveillance-driven AI should be treated as a political, not merely technical, issue

Because AI depends on large-scale data extraction, its deployment implicates citizenship, privacy, and power. The practical implication is that societies must draw principled limits around behavioral monitoring and predictive governance, rather than treating such systems as neutral efficiency gains. On the one hand, this requires instrumental intervention to prevent the intrusion of universal "datafication". On the other hand, it requires also the continual cultivation of virtue, as antidote to the deeper root from which technological rationality springs—namely, the will to turn one's gaze from truth to power.

II. Descriptive Analysis

The seminar's consensus on "artificial intelligence" is neither that AI is a neutral instrument nor that it is an autonomous super-agent destined to replace human beings. Rather, AI is best understood as a particularly powerful and revealing expression of modern *technique*: a system of prediction, control, and optimization that intensifies tendencies already present in technological modernity. The current trajectories of its development—the particular forms it is taking—increasingly render it a harmful technology.

First, AI belongs to what is described as a "universal engineering" mentality. It operates by reducing complex realities to manipulable variables and then reordering them toward efficiency and control. This is precisely why AI "works" so well—it not only has a mechanism for converting language and manipulable variables, but operates within the widespread accepted framework of "universal engineering". Yet that same success gives it the misleading appearance of cultural and intellectual authority, encouraging the assumption that what can be modeled, predicted, and optimized exhausts what is real or worthy of human attention, while other modes of understanding—teleological, moral, contemplative—are displaced and marginalized, being unsuited to this "universal engineering" framework.

Second, claims about intelligence "emerging" from AI systems must be treated with extreme skepticism, for two reasons. First, appeals to emergence are often used as substitutes for explanation, suggesting that scale and complexity alone can produce understanding or judgment. While complexity may explain certain dimensions of performance within circumscribed systems, it does not explain the constitution of those systems themselves. Second, such claims often presuppose an unstated definition of "intelligence", which runs the risk of *equivocation*. That is, intelligence in the proper human sense—which involves

truth, meaning, and purposive judgment inexorable from the experience of a living knower—includes but does not reduce to the capacity for problem solving or accurate association; and thus if we reduce "intelligence" to this lattermost capacity and predicate it equally of man and machine, we err by equivocation (mediated by the fallacy of *pars pro toto*).

Third, AI is recognized as inseparable from surveillance and data-extraction economies. Its effectiveness, especially at increasing scales of operation, depends upon comprehensive behavioral data, leading to the construction of detailed predictive models of individuals and populations. The concern here extends beyond advertising or convenience to the broader concentration of social and political power that follows from large-scale behavioral prediction and manipulation. Moreover, this comprise of all things within discrete data-points enabling successful AI-implementations results in a kind of self-fulfilling prophecy, wherein rebellion takes the form of unpredictability, as the implementations reward conformity to pre-existent patterns.

Fourth, the deepest danger of AI is not primarily erroneous or biased outputs—which may indeed cause extensive and intense proximate damages—but rather the way it integrates into existing systems so as to gradually absorb (or dissolve) human agency. By coupling AI to infrastructure—logistics, medicine, finance, governance—decision-making shifts away from the exercise of prudential judgment toward rapid and opaque automated "optimization". Human ends may still be named, but the means by which they are pursued increasingly determine the outcome, often without clear responsibility or oversight or even the possibility of prudential reservation, questioning, revision, and restriction. Far more dangerously, when it integrates into the technologies closer to human formation, such as reading, writing, communicating, socializing, and so on, here we have the potential for automated optimization occurring simultaneously with a complete and total atrophy of all possibility for virtuous human development. Rather than autonomous agents ourselves, we become radically dependent upon the machines, and therefore radically vulnerable in ourselves, lacking all moral and psychological resilience.

Fifth and finally, we must recognize the possibility of concrete patterns in which human life itself becomes organized as an algorithmic regimen. Individuals set goals, then cede increasing control of daily decisions to systems that promise optimization and efficiency. While such arrangements may succeed by certain metrics—perhaps even all the metrics we might establish—they exemplify a deeper risk: the gradual habituation of persons to live under technical governance (i.e., "good" being determined *by* those metrics themselves) rather than personal judgment—a modality of life which can be most effectively totalized under the advancing tendencies of LLM development.

In sum, we advocate a view of AI as **psychologically** and **politically non-innocent**. As presently construed, it accelerates the tendency of technological systems to reconstitute human action under the signs of efficiency and control. The implied response is not wholesale rejection, but a radical vigilance: preserving domains of human responsibility—truth-seeking, prudence, moral agency—that cannot be delegated to machines without deforming both the person and the polity.