The Intellectual Agenda of AI Humanism

A Research Program for Human Capability in the Age of Artificial Intelligence

I. Purpose

Artificial intelligence expands the range of what humanity can know, create, analyze, and accomplish.

AI Humanism studies what happens next.

Its central question is:

What can human beings become in an age of capable machines?

The Intellectual Agenda of AI Humanism translates that question into a program of scholarly inquiry.

The Agenda studies how artificial intelligence changes the development and exercise of human capability; how institutions organize human and machine intelligence; how abundant information becomes durable knowledge; and how delegated machine execution remains connected to human purpose.

It organizes this inquiry around the four pillars of AI Humanism:

Human Agency
Institutional Integrity
Knowledge Stewardship
Governed Intelligence

These pillars operate at different but connected levels. Human Agency concerns people's capabilities. Institutional Integrity concerns the organization of collective action. Knowledge Stewardship concerns the development and preservation of understanding. Governed Intelligence concerns the mechanisms through which institutions translate purposes into machine-enabled execution.

Together, they support a common research program: understanding how advances in machine capability can become advances in human and institutional capability.

The Agenda is provisional by design. Its questions should evolve as evidence accumulates, technologies change, and scholarship exposes better questions.

Its ambition is durable: to develop a body of knowledge about human development under conditions of increasingly abundant machine intelligence.

II. Human Capability as an Object of Study

AI Humanism treats human capability as something that develops.

People acquire expertise. They learn to reason within disciplines. They develop judgment through experience. They learn to communicate, create, coordinate, interpret evidence, formulate purposes, and act within institutions.

Institutions also develop capabilities. They accumulate knowledge, organize expertise, establish authority, coordinate action, learn from experience, and become better or worse at accomplishing their purposes.

Artificial intelligence enters these developmental processes as a new source of cognitive and executable capability.

This creates the central empirical problem of AI Humanism.

The Capability-Conversion Problem

Under what conditions does an increase in machine capability produce a durable increase in human or institutional capability?

This question rests on a foundational analytical distinction:

Machine capability, assisted performance, and human or institutional capability should be measured as distinct constructs.

They may move together, but they need not.

Machine capability concerns what an intelligent system can do.

Assisted performance concerns what a person or institution can accomplish while using that system.

Human or institutional capability concerns what remains as a durable capacity to understand, judge, create, coordinate, learn, decide, and act.

A more capable system may improve assisted performance immediately without producing a corresponding increase in durable human or institutional capability. In other settings, the same technology may accelerate expertise formation, improve judgment, strengthen organizational learning, or expand the range of work an institution can reliably undertake.

AI Humanism therefore studies more than whether AI improves an outcome.

It asks what changes in the person, institution, or knowledge system that produced the outcome.

Human capability includes the durable capacity to understand, judge, formulate problems, create, communicate, coordinate, decide, and act. Institutional capability includes the capacity to organize expertise, allocate authority, preserve knowledge, learn, adapt, and execute toward a purpose.

These capabilities can occur at different levels and be measured through different disciplines. Their development over time is a central object of AI Humanist scholarship.

This orientation gives the field a distinctive unit of analysis:

The relationship between machine capability, assisted performance, and the development of human and institutional capability.

III. Human Agency

Guiding Principle

Delegate execution while preserving human authorship.

Human Agency studies how artificial intelligence changes the range over which people can form purposes, exercise judgment, make consequential choices, and act.

AI expands access to information, analysis, expression, experimentation, and execution. These capabilities can allow individuals to undertake work previously requiring larger teams, specialized intermediaries, or substantial organizational resources.

The research question is therefore larger than whether people become more productive.

It asks whether people become more capable.

Research Domains

A first domain concerns **judgment**. As machines perform more analysis and generate more plausible alternatives, research should examine how people learn to evaluate evidence, compare alternatives, recognize uncertainty, and determine what deserves attention.

A second concerns **problem formation**. Abundant answers may increase the value of identifying worthwhile questions, framing problems correctly, specifying purposes, and recognizing when the initial formulation is inadequate.

A third concerns **authorship and delegation**. People increasingly direct work that they do not personally execute. Research should examine how meaningful authorship operates when execution is distributed across human and machine contributors.

A fourth concerns **expertise formation**. AI can provide explanation, simulation, feedback, and access to specialized knowledge. It can also perform activities through which expertise has historically developed. Research should determine how different forms of AI use affect learning, retention, transfer, and independent performance.

A fifth concerns **the scale of individual agency**. AI may allow individuals to coordinate resources, analyze information, create products, conduct research, or operate enterprises at scales previously available primarily to organizations. This expansion deserves study as a change in the effective capabilities of the individual.

Initial Research Questions

- When does AI-assisted performance produce transferable human capability?
- How does repeated AI use affect judgment, problem formulation, and independent reasoning over time?
- Which forms of delegation preserve meaningful human authorship?
- How does AI alter the path through which novices become experts?
- Which capabilities become more valuable as machine execution becomes more capable?

Human Agency research should distinguish what a person can accomplish **with** a machine from what the person becomes capable of understanding and doing **because of** sustained interaction with one.

That distinction is central to the developmental project of AI Humanism.

IV. Institutional Integrity

Guiding Principle

Execution can move. Accountability remains.

Human beings extend their capabilities through institutions.

Artificial intelligence changes how institutions distribute analysis, expertise, communication, coordination, and execution. Work can move between employees, experts, software systems, external providers, and autonomous or semi-autonomous agents.

Institutional Integrity studies how organizations develop under these conditions.

Its primary unit of analysis is the organization of **purpose, authority, responsibility, and accountability** around machine-enabled work.

Research Domains

A first domain concerns **authority**. Institutions need ways to determine who may delegate what, which systems may act, how far delegated authority extends, and when human authorization remains necessary.

A second concerns **accountability**. Machine participation changes execution without eliminating the institutional requirement to identify who remains responsible for consequential action.

A third concerns **organizational design**. AI may change the value of hierarchy, specialization, coordination, managerial attention, professional expertise, and firm boundaries. Research should examine which organizational forms become more effective as intelligent execution becomes increasingly available.

A fourth concerns **expertise and decision rights**. Institutions traditionally distribute authority partly according to specialized knowledge. AI changes the distribution and cost of access to that knowledge. Research should examine how expertise, judgment, and decision authority should relate when sophisticated analysis becomes broadly accessible.

A fifth concerns **institutional learning**. Organizations improve when experience becomes knowledge that can inform future action. Machine-mediated work creates new forms of evidence while also changing how experience is generated and retained.

Initial Research Questions

- How should institutions allocate authority when machines participate directly in execution?
- Which forms of accountability remain effective when execution becomes distributed?
- How does AI change the relationship between expertise and decision rights?
- Which organizational structures best convert machine capability into institutional capability?
- How can institutions learn from machine-enabled execution rather than merely repeat it?

Institutional Integrity treats the institution as an active site of human development.

Organizations do more than consume technology. They determine how capability is coordinated, accumulated, directed, and preserved.

V. Knowledge Stewardship

Guiding Principle

Use intelligent machines to expand what humanity knows and deepen humanity's capacity to understand.

Artificial intelligence makes information and sophisticated analysis increasingly abundant.

Knowledge remains a developmental achievement.

Knowledge Stewardship studies how people and institutions convert informational abundance into understanding, expertise, discovery, and durable intellectual capability.

Research Domains

A first domain concerns **abundance and judgment**. When retrieval, synthesis, and fluent explanation become inexpensive, intellectual scarcity may move toward question selection, interpretation, verification, causal reasoning, and judgment.

A second concerns **expertise**. Expert knowledge includes more than access to correct propositions. It includes pattern recognition, contextual understanding, tacit knowledge, disciplinary standards, and the ability to recognize when ordinary rules do not apply. Research should examine how AI changes the formation and exercise of these capabilities.

A third concerns **knowledge creation**. AI can search larger spaces of ideas, identify patterns, propose hypotheses, synthesize literature, and assist experimentation. Research should examine where these capabilities genuinely expand discovery and how human judgment contributes to that process.

A fourth concerns **knowledge preservation**. Institutions depend on intellectual memory. As people increasingly retrieve answers through machine intermediaries, research should examine how knowledge is retained, documented, transferred, and reconstructed.

A fifth concerns **participation in knowledge production**. Lower barriers to analysis and expression may allow more people to participate in sophisticated scholarship, science, professional work, and creative practice.

Initial Research Questions

- How does abundant machine synthesis change the value and formation of human expertise?
- Which forms of knowledge become more durable through AI, and which become more fragile?
- How does AI affect the ability to distinguish information, explanation, evidence, and understanding?
- Under what conditions does AI accelerate genuine discovery rather than the recombination of existing knowledge?
- Can AI broaden participation in knowledge creation while strengthening standards of inquiry?

Knowledge Stewardship therefore examines both the frontier of knowledge and the human capacity to reach it.

The objective is an expanding body of human knowledge joined to an expanding capacity for human understanding.

VI. Governed Intelligence

Guiding Principle

Delegated intelligence should remain directed, observable, and revisable.

Artificial intelligence increasingly translates human purposes into action.

Governed Intelligence studies the mechanisms that connect those purposes to machine execution.

It begins where institutional authority becomes operational: a purpose becomes a mandate, authority is delegated, a system acts, evidence records the execution, and people or institutions assess whether action remains aligned with the original purpose.

This pillar therefore studies the architecture of intelligent delegation.

Research Domains

A first domain concerns **mandates**. Human purposes often contain objectives, constraints, acceptable sources, evidentiary requirements, escalation conditions, and contextual judgments. Research should examine how institutions represent these elements in forms that can meaningfully direct machine execution.

A second concerns **bounded authority**. Delegation requires a relationship between the capabilities a system possesses and the authority it has been granted. Research should examine how these boundaries can remain intelligible as systems become more capable and workflows more dynamic.

A third concerns **evidence of execution**. Accountability requires knowledge of how delegated work occurred. Research should identify which forms of process evidence allow people to understand whether machine execution remained faithful to its mandate.

A fourth concerns **correction**. Governed systems should allow institutions to identify deviations, intervene when necessary, revise mandates, and incorporate experience into subsequent execution.

A fifth concerns **adaptive governance**. Different forms of work require different levels of direction, evidence, escalation, and review. Research should examine how governance can vary with uncertainty, consequence, reversibility, and institutional context.

Initial Research Questions

- How can institutions represent human purposes in operational mandates without removing the judgment those purposes require?
- What forms of execution evidence support meaningful review and accountability?

- How should delegated authority change with task uncertainty, consequence, and reversibility?
- Where within machine execution are different forms of direction, observation, and correction most effective?
- How can institutions expand delegation while preserving the capacity to understand and revise what intelligent systems do?

Governed Intelligence connects capability with institutional purpose.

Its research objective is the development of mechanisms that allow increasingly capable systems to participate in increasingly consequential work while remaining meaningfully directed by human institutions.

VII. Human Formation and the Scholar-Citizen

AI Humanism joins scholarship with human formation.

The four pillars therefore converge on an educational question:

What should an educated and capable person be able to understand and do in an age of intelligent machines?

This question extends beyond technological literacy.

The AI age increases the importance of reasoning independently, evaluating evidence, interpreting generated language, formulating consequential questions, communicating clearly, understanding institutions, and exercising judgment when competent analysis is readily available.

Rhetoric becomes especially significant.

Generative systems can produce fluent language at extraordinary scale. Human capability therefore increasingly depends on understanding the relationships among fluency, evidence, argument, persuasion, interpretation, and meaning.

Historical understanding also matters. Artificial intelligence is consequential, but humanity has experienced earlier transformations in knowledge, communication, production, organization, and professional expertise. Historical comparison can distinguish genuinely new phenomena from recurring institutional patterns.

AI Humanism carries the ideal of the scholar-citizen into this environment: a person who seeks understanding and brings that understanding into professional, institutional, intellectual, and civic life.

Research on human formation should ask:

- Which intellectual capabilities increase in value as machine intelligence becomes abundant?
- Which educational practices best develop independent judgment in AI-rich environments?
- How should rhetoric and interpretation change when machines generate persuasive language?
- What forms of technological understanding allow non-specialists to exercise meaningful judgment about AI-enabled work?
- How can education use AI to expand access to sophisticated intellectual activity while developing durable expertise?
- What should distinguish an educated AI user from a merely proficient one?

The findings of this research program should inform **Studia Humanitatis**, the educational framework of AI Humanism.

The research must precede prescription. The educational program should develop from evidence about the capabilities people need and the practices through which those capabilities actually form.

VIII. Cross-Pillar Research Programs

The four pillars identify different levels of inquiry, but many of the most consequential questions occur between them.

AI Humanism will therefore develop research programs that cross pillar boundaries.

1. Capability Conversion

The first asks when machine capability becomes human capability.

A productive system may improve immediate performance. A developmental system may also change what its user can subsequently understand, judge, create, or accomplish.

Research should identify the mechanisms that distinguish these outcomes and determine how they vary across tasks, people, institutions, and time.

2. Delegation and Development

Delegation changes both work and the people performing it.

When machines assume execution, people may redirect attention toward problem formulation, interpretation, coordination, judgment, or new forms of creation. The resulting division of cognition and execution can alter how expertise develops and how responsibility is understood.

Research should examine delegation as a process of human and institutional development, not simply task allocation.

3. Abundance and Scarcity

AI can make some intellectual resources abundant.

When information, synthesis, drafting, coding, or routine analysis become less scarce, other capabilities may become more consequential. These may include judgment, attention, problem selection, verification, contextual understanding, institutional knowledge, creativity, and the capacity to establish purposes.

Research should identify how the structure of intellectual scarcity changes as machine capability expands.

4. Institutional Amplification

The same technological capability can produce different results in different institutions.

Organizational structure, incentives, authority, expertise, evidence, culture, and learning processes may determine how effectively machine capability converts into institutional capability.

Research should identify these mediating conditions and explain why similar technologies produce different developmental trajectories across organizations.

5. Capability Across Time

Many effects of artificial intelligence will be cumulative.

Short-term productivity measures may reveal little about expertise formation, institutional memory, intellectual independence, or adaptive capacity over years.

AI Humanism therefore places particular value on longitudinal research that examines what individuals and institutions become through sustained use of intelligent systems.

IX. Initial Research Propositions

A research agenda should expose its assumptions to examination.

The following propositions establish theoretical expectations for investigation. They invite empirical testing, refinement, and competing explanations.

Proposition 1: Immediate performance and durable development may diverge across time.

AI may improve an immediate output while producing different long-term effects on expertise, judgment, learning, and institutional knowledge.

Research should therefore examine both performance and development.

Proposition 2: As execution becomes easier to delegate, the relative importance of purpose and judgment may increase.

Machine capability can shift human contribution toward determining objectives, framing problems, interpreting evidence, evaluating alternatives, and revising direction.

Research should identify **whether and where** this shift occurs.

Proposition 3: Institutional arrangements mediate the conversion of machine capability into institutional capability.

The effects of AI depend partly on how institutions allocate authority, preserve accountability, organize expertise, learn from experience, and govern execution.

Research should examine which institutional arrangements strengthen or weaken capability conversion.

Proposition 4: Informational abundance shifts the locus of intellectual scarcity.

As access to information and analysis expands, scarcity may move toward judgment, verification, interpretation, attention, contextual knowledge, and question formulation.

Research should determine where scarcity moves, how strongly it shifts, and which capabilities become more consequential in different settings.

Proposition 5: Greater governability can expand the feasible domain of delegation.

Institutions may delegate more consequential work when they can direct execution, observe relevant evidence, identify deviations, and revise action.

The feasible domain of delegation will also depend on consequence, legitimacy, economics, and institutional context.

These propositions give the Agenda a starting theoretical structure.

Their value lies in making the expectations of AI Humanism visible, contestable, and researchable.

X. Methods and Evidence

AI Humanism is interdisciplinary because its questions cross levels of analysis.

Method should follow the question.

Research may draw from controlled experiments, longitudinal studies, organizational field research, comparative case studies, historical analysis, ethnography, surveys, computational methods, technical prototyping, design science, economics, psychology, organizational theory, education research, philosophy, and the humanities.

Interdisciplinary breadth requires disciplinary seriousness.

AI Humanism therefore adopts several methodological priorities.

Measure Capability, Not Only Performance

Research should distinguish immediate task outcomes from changes in durable human or institutional capability whenever the distinction is relevant.

A productivity increase answers one question. Whether the user learned, retained expertise, improved judgment, or expanded independent capability answers another.

Both matter.

Study Development Over Time

Many important consequences of AI use emerge through repeated interaction.

Longitudinal research should therefore examine learning, expertise, institutional memory, dependency, adaptation, changing authority, and capability accumulation.

Examine Multiple Levels

Individual behavior occurs within teams, organizations, professions, and institutions.

Research should identify the relevant unit of analysis and examine cross-level effects when they matter. An intervention that benefits an individual user may produce different consequences for an organization, and organizational changes may alter the development of individual capability.

Study Execution as Well as Outcomes

Machine-generated outputs provide incomplete evidence about machine-enabled work.

Research into delegation and governance should examine the processes, decisions, evidence, and interventions through which outcomes are produced.

Connect Scholarship with Practice

Institutions provide settings in which to examine theories about AI-enabled work under real operating conditions.

The Institute should cultivate research partnerships that enable observation, experimentation, comparison, and longitudinal study while preserving scholarly independence.

Preserve Contestability

Research should make assumptions visible, distinguish evidence from interpretation, document relevant limitations, and invite competing explanations.

AI Humanism advances when better evidence changes its conclusions.

XI. A Cumulative Research Program

The Intellectual Agenda of AI Humanism is intended to produce cumulative knowledge.

Individual studies should contribute to larger questions about capability formation.

Research on judgment should help explain Human Agency. Research on organizational authority should develop Institutional Integrity. Research on expertise should deepen Knowledge Stewardship. Research on machine delegation should advance Governed Intelligence.

The strongest work will connect these levels.

A study of AI-assisted professional decision-making, for example, might examine immediate task performance, changes in individual expertise, changes in organizational decision rights, the preservation of institutional knowledge, and the evidence required to govern delegated execution.

Such work would reveal more than whether the technology worked.

It would show how machine capability changed the capabilities of the people and the institution.

The Institute should develop theories, empirical studies, measurement frameworks, comparative cases, technical experiments, educational research, and practical methods that allow these questions to accumulate into a recognizable body of scholarship.

Over time, the Agenda should make it possible to answer increasingly precise versions of the question at the heart of AI Humanism:

When does artificial intelligence enlarge human agency?

When does it strengthen institutions?

When does informational abundance become knowledge?

When does delegated intelligence become governable capability?

And ultimately:

When does the advancement of machine capability become an engine for the advancement of human and institutional capability?

That is the research program of AI Humanism.

The humanist project continues through inquiry.