



THE BEACON PAPERS • No. 15

THE HUMAN ADVANCEMENT TEST

How Human–AI Co-Creation Advances Humanity

Nguyen Anh Tuan

Founder and Chief Architect of AIWS
Co-Founder, Co-Chair and CEO, Boston Global Forum

AIWS Lumina Lab • Beacon Hill, Boston • Columbus Day, 12 October 2026

Better output does not, by itself, demonstrate human advancement.

I. Ten Minutes

In randomized controlled trials published in 2026, researchers from Carnegie Mellon University, the University of Oxford, MIT, and UCLA gave 1,222 people problems in mathematical reasoning and reading comprehension. With an AI assistant, people performed better. When the assistant was taken away, they performed significantly worse than people who had never used it, and they were more likely to give up. The effect appeared after about ten minutes.

Ten minutes is one conversation. The finding concerns particular tasks and cannot be stretched to all work. But it shows that the work in front of a person can improve while the person's own capacity weakens, and that nothing in the work reveals it. Any school, newsroom, hospital, firm, or government that adopts AI therefore has reason to measure what happens to its people, not only to its output.

In the AI Age, every person should have the chance to grow in intelligence, to raise their creative power, and to be the master of their own future. The results of working with AI must therefore be judged on two counts: the value created for life, and the growth of the human being who created it. A student completes good research with AI. A journalist finds a new angle. A team builds a useful service. Afterward: what do they understand more deeply? Can they judge harder questions, develop ideas of their own, and take fuller responsibility?

The Human Advancement Test (HAT) puts these questions into practice. **HAT helps people advance their intelligence throughout life, with the help of AI Assistants.** Beacon Papers

No. 8 introduced it as the measure of AIWS Civilization and warned that a test that cannot be failed is not a test. Beacon Papers No. 7 called for instruments and stated that none was yet operational. Beacon Papers No. 14 made the Test one of the four conditions of Human–AI Co-Creation. **This paper is the official development of the Test.** It sets out its dimensions, its method, and its rules of evidence, and it supersedes the five-part division of Beacon Papers No. 8.

The question set and the standard of the Human Advancement Test originated at AIWS Lumina Lab, on the foundation of Love · Creativity · Nobility · Wisdom. Their lasting aim is that every person can say clearly: where I have advanced, what proves it, and how I will continue to grow.

II. Two Records, Never One

Every HAT record has two parts that travel together and are never merged: **the value of the work**, and **the advancement of the person who did it**. The quality of the product, its benefit to users, and its real-world impact are recorded on their own terms. Evidence of the person’s growing capacity is recorded separately.

The separation is the heart of the Test. A useful result does not by itself prove that its maker understands more. The study shows why: output and capacity can move in opposite directions, and while the work is under way only the output is visible. This is the co-production trap named in Beacon Papers No. 14. A person or institution that judges AI only by its results may keep reporting success without recognizing a decline in its people’s ability to think, persist, and decide on their own. The reverse also holds: a meaningful experience of learning may need more time before it becomes a contribution to others. Both deserve to be seen, and neither may stand in for the other.

III. Three Readings

HAT recognizes human advancement through two kinds of capacity: independent capacity, and purposeful co-creative capacity. The quality of the product, by itself, proves neither. Independent capacity is what the person can understand, explain, and apply on their own. Co-creative capacity is what the person contributes within the collaboration: the questions they pose, the goals they choose, the standard they hold the AI to, the process they lead, and the care they bring. Both are human capacities, and both are measured against the person’s own starting point.

The Test therefore reads every piece of work at three points:

Before — the starting point: what the person could already do, kept as a product or an explanation. Where no such record exists, the starting point is self-described and noted as less certain.

With — what the person achieved with AI. The product itself is the value of the work, not evidence of advancement. What the With reading can evidence is the person’s own contribution within the collaboration — a question, a choice, a standard, a direction — when it can be identified and compared with the starting point.

Without — what the person can do once AI's help with content is withdrawn: explain a decision in their own words, solve a variant of the problem, apply what was learned to a new task.

From these readings come three kinds of claim, each worth recording with its own evidence and level of support: **done with AI; understood, and able to be explained and verified; applied to a new situation.** The first records the value of the work; the second and the third, together with the person's identified contribution within the collaboration, are claims of human advancement.

The Without check is the decisive test of understanding and independent application. It is not the only door to evidence. The ability to lead AI is itself a human capacity, and it can only be shown in the interaction. Love for people and nature cannot be confirmed by taking the AI away.

The Without check removes AI's help with content, never a person's means of access. Supports a person needs in order to take part — for disability, language, or circumstance — stay in place, and the conditions are recorded.

The principal comparison is the person with themselves over time, in their own circumstances. People begin from different starting points, conditions of learning, access to technology, and aspirations. HAT measures the distance each person travels, not their rank against others, and it must widen the opportunity to grow for everyone who takes part.

IV. The Six Dimensions

This paper establishes six dimensions of human advancement as the official form of the Test. They develop and replace the five named in Beacon Papers No. 8. Human Intelligence becomes two dimensions that can be evidenced separately: understanding and judgment. Human Agency becomes command. Human Creativity remains. Human Dignity becomes the guarantee of the whole Test rather than one of its scores: no person's dignity depends on a HAT result. Human Wisdom lives in judgment and in love. And two dimensions are added that No. 8 did not contain — the person's power to raise the quality of the AI, and the person's love for people and nature.

1. Deeper Understanding

What do I now understand, and how can I explain it in my own words? Evidence: a misconception corrected, a new connection between fields, the ability to use knowledge in a different situation. Repeating a correct answer from AI does not show that understanding has formed.

2. Sounder Judgment

What important choice did I make, and on what grounds? The person records an AI proposal that was verified, adjusted, accepted, or rejected, and why. Sound judgment also shows in recognizing what one does not know, finding the right person to ask, and changing one's view when the evidence changes.

3. Higher Creativity

What question, idea, or approach of real significance did I contribute? The contribution may lie in defining the right problem, combining sources of understanding, designing an experiment, or developing a new direction through dialogue with AI. Drafts and decisions to change course make

the process visible. The person distinguishes what is new to themselves from a claim of novelty in a field.

4. Fuller Command

How far can I lead the work, and how do I answer for it? Evidence: setting the purpose, assigning AI suitable tasks, detecting its limits, seeking expert help, and correcting errors. Command includes the right to stop, to change course, or to refuse a proposal that lacks sufficient grounds. It is Human Agency made concrete (Beacon Papers No. 5 and No. 11).

5. Elevating the AI Assistant

Where did I see that the AI's suggestion fell short — inaccurate, shallow, or unsuited to the situation? What higher standard did I require? After the AI responded, on what grounds do I judge that the AI Assistant served better?

The person keeps one before-and-after example: the first suggestion, what needed fixing, the request to raise it, the revised response, and their own assessment. Improvement may show in accuracy, depth, creativity, fitness to circumstance, or care for people. If the AI still falls short, the record states what remains unmet and the next check.

This capacity can only be shown within the collaboration. A person who can raise the AI holds a standard the AI did not supply. It is the human side of Co-Elevation. What is recorded is improved assistance in the work under review; it does not prove that the model has learned permanently or performs better everywhere.

6. Greater Love for People and Nature

After each result of working with my AI Assistant, do I love people and nature more? Whose circumstances do I now understand better? Have I come to care more for life, or changed a choice in order to care for, protect, or avoid harm?

The person may record an insight, a feeling, or a concrete action: listening to those affected, correcting disrespectful wording, weighing the effect on nature, or doing a practical act of care. What is felt, what was done, and what effect was observed are recorded separately. "No change noticed yet" is an honest answer. Love must never be forced into an assertion or a score.

Across all six

Running through every dimension is one question: whom does the work help, what effects might it have, and who bears final responsibility? Responsibility must rest with a person or an organization able to answer and to act (Beacon Papers No. 10).

A single project may show clear advancement in one or two dimensions. The record states exactly that scope and shows that command and responsibility were maintained. No piece of work is required to show equal growth in all six.

V. The Test Judges the Use, Not the Tool

The study's most important finding lies in a detail. In its analysis by self-reported use, the decline was concentrated among participants who asked AI for direct answers, while those who used it only

for hints or clarification performed close to participants who never had AI. The authors state that this comparison is not necessarily causal. Still, it shows results differing with the way the same assistant was used on the same task.

HAT therefore does not certify technologies as good or bad. It judges the relationship between a person and a system: how AI is used, for what, and with what effect on the human being. A school that uses an assistant to give answers and a school that uses the same assistant to ask better questions are building two different futures.

This places an obligation on those who design AI: **build for the Without reading**. Systems should leave the generative step to the human, expose reasoning rather than only results, preserve productive difficulty, and support unaided practice — the conditions of Co-Elevation set out in Beacon Papers No. 8.

VI. The Test in One Piece of Work

Before beginning

Choose work that is meaningful and concrete enough to finish. Write down briefly the problem to be solved, who may benefit, what you can already do, and one capacity you want to develop. Keep an initial product or explanation as the Before reading. Decide the intended role of AI and the decisions that remain human: AI may help find lines of analysis, while the person chooses the question, checks the sources, and defends the conclusion.

During the work

Keep a few moments that show your judgment and contribution: the problem you set, the proposal you received, the decision you made, and why. Keep only the essential traces — drafts, sources checked, notes from trials, feedback. There is no need to keep the whole conversation with AI.

On completion

Compare the result with the original goal. Answer the six questions, tying each answer to evidence or concrete experience and separating self-perception from what has been checked. Record separately the value of the product, who benefits, its limits, and who is responsible; benefits not yet observed are recorded as expectations.

The Without check

With AI's help on content withdrawn, explain a decision in your own words, handle a variant of the problem, or apply what you learned to a new task. For understanding and independent application, this is the reading that turns a claim into evidence.

The verdict and the next step

For each dimension, conclude in one of three ways: **advancement evidenced; not yet evidenced; or signs of decline — the practice must change**. The third verdict is what makes HAT a test: it must be able to return an unwelcome answer. A person who cannot yet explain their own product returns to a key decision, checks its grounds, and tries to apply it again — and changes how they use AI, from asking for answers to asking for hints, explanations, and challenges.

Conclusions are specific: “I can now check sources on my own and found one error; I still need help building a causal argument.” Narrow evidence permits only a narrow conclusion. The next goal should be within reach but more demanding.

VII. Growth Over the Years

Each piece of work creates a HAT milestone. Over time, milestones form a record of a person’s intellectual development: the capacities already held, what is newly formed, the difficulties that remain, and the direction of further growth. The record keeps corrections and lessons learned from mistakes. It also shows a growing ability to guide AI, and a growing love for people and nature, through insight, choice, and action.

A sustainable rhythm: after each project, record a milestone; each quarter, review a few that matter; each year, answer three questions — Where have I grown? What proves it? What capacity will I develop next? Each person adapts this rhythm to their work and circumstances.

Comparison must be fair. A better product may come from a stronger AI, more time, or a colleague’s help. Milestones therefore record the difficulty of the task, the tools and level of support, the roles of others, and the conditions of the work. Where possible, comparable tasks or a common application exercise are used. Where conditions differ too much, the change is described but no gain in capacity is claimed. Capacity retained and applied after an interval is further evidence of growth. A single milestone cannot show lasting progress or prove that AI was the only cause. Review by a mentor, a colleague, or a user checks self-assessment.

An illustration, not a required path: a person begins by using AI to summarize documents but cannot yet separate claim from evidence. At the next milestone, they check sources on their own and correct an unsupported conclusion. Later, they pose new questions, design the check, and defend their choices. Further on, they guide others in similar work and explain the limits of the method. Each step must be shown in real work. After each review, the person chooses one priority capacity, one piece of work to practice it, and one way to check the result. HAT has value when looking back leads to the next practice.

VIII. Three Examples

The cases below are illustrations, not results of HAT trials.

A student researching a local problem. The aim is to understand a local environmental problem and propose a survey. AI suggests questions and a report structure; the student sets the scope, reads primary sources, and explains the method. A fluent report the student cannot defend is not evidence of advancement. A milestone might record that the student found a source that did not exist, narrowed a question that was too broad, and applied the method unaided to another locality. The value of the proposal to the community is checked separately.

A journalist developing a story. The aim is to help readers understand a social problem. AI organizes material and suggests lines of inquiry; the journalist verifies, interviews, chooses the angle, and answers for the edit. Faster publication is recorded as a result of the work. HAT

asks for evidence of professional capacity: a contradiction found in the documents, an interview question that changed the story, an unsupported inference corrected — and, at the next review, whether the same method of verification holds on a new topic. Sources are protected when evidence is shared.

A team improving a service. The aim is to reduce difficulty for users. AI summarizes feedback and proposes solutions; the team chooses, tests, and checks the effect, recorded with real feedback or data over a stated period. Each member keeps their own record. The team's success does not prove that every member advanced equally: one may understand users more deeply, another may design better experiments. The next step is a new task, suited to each person, that checks what they can apply.

In each case HAT keeps the link between real contribution and personal growth. Difficulties recorded honestly become the basis for changing how one learns and how one works with AI.

IX. Three Levels of Practice

The person. Anyone can take the Test, alone, with the template that follows this paper. It is the practice behind the third commitment of the AIWS Human–AI Co-Creation Movement: *I use AI to become more capable, never less.*

The Elevate Circle. The co-production trap is hard to see from inside. Elevate Circles, the local groups of the Movement, take the Test together: members set one another's application tasks, witness one another's Without checks, and review one another's records. Witnessing turns a private habit into evidence.

The institution. Schools, universities, hospitals, enterprises, and public agencies apply the Test across their people with the methods Beacon Papers No. 8 requires: longitudinal readings, comparison groups, methods stated in advance, and publication of negative results on the same terms as positive ones. Independently assessed, institutional results are the basis on which the criteria of AIWS Co-Elevation Certified are to be developed. Certification applies to institutions; persons are never graded by HAT.

X. Rules That Keep the Test Honest and Human

- 1. It must be able to fail.** A test that always passes measures nothing.
- 2. The product is not the proof.** No output achieved with AI counts, by itself, as evidence of human advancement. Evidence comes from independent capacity shown without AI's help on content, and from the person's identified contribution within the collaboration — each compared with the starting point.
- 3. Commitment is not certification.** Joining the Movement is a promise; passing the Test is a demonstration (Beacon Papers No. 14).
- 4. Declines are reported.** A decline found in time is the most valuable result HAT produces. AIWS has committed to publishing such findings (Beacon Papers No. 8).

5. **The record belongs to the person.** The participant keeps the record and may review, add to, and correct it, and chooses what to share. A record of advancement must never become a fixed label on a human being. No person's dignity depends on a HAT result.
6. **AI's praise is not evidence.** AI may help pose questions, organize evidence, and spot inconsistencies, but praise or assessments generated by AI prove nothing on their own. Where greater reliability is needed, a mentor or colleague reviews the evidence, and the scope of that review is stated.
7. **Description before scores.** The practice uses grounded description, not a composite score. Any future scale must first be studied for reliability, fitness across fields, the effect of language, and conditions of access. Until it is validated for such purposes, HAT is never the sole basis for hiring, academic grading, or the allocation of opportunity.

XI. HAT in AIWS Civilization

The AIWS Human–AI Co-Creation Movement brings HAT into concrete work in schools, enterprises, and communities. Each person chooses work worth doing, does it with AI, and records what they contributed and what grew in themselves. Mentors turn the evidence into feedback for the next step.

The Human–AI Co-Creation Atlas will receive the records participants choose to share: the contribution, the roles of human and AI, the evidence of advancement, and who is responsible. A public record distinguishes self-assessment from assessment reviewed by others. The full personal record can remain private; only what is necessary and consented to is published. Shared cases show practices others can learn from. The number of records and success stories does not by itself show that a society has advanced; claims at that scale require research with suitable methods and scope.

The next step can be a voluntary pilot across many fields, examining ease of use, the quality of evidence, and how well HAT helps participants choose their next step of learning. Its results will refine the method before it is widened.

A civilization measures what it values. The industrial age measured output. The AI Age has so far measured the intelligence of its machines. AIWS Civilization adds the measure that decides whether the others matter: what human beings understand, judge, create, command, raise, and love — with AI and without it.

HAT helps people advance their intelligence throughout life, with the help of AI Assistants. AIWS Civilization aims at people who grow ever more knowledgeable, creative, self-directed, loving, and responsible. From AIWS Lumina Lab, the Human Advancement Test carries that aspiration into practice, year after year: do meaningful work, recognize what has grown, raise the quality of collaboration with AI, and keep growing in order to serve people and nature.

BEACON PRINCIPLE No. 15

Measure the person, not the output.

*Test what grows, with AI and without it.
Record what has grown.*

AI ADVANCES. HUMANITY MUST ADVANCE FURTHER.

The Human Advancement Test · Practice Template

The question set and the standard of the Human Advancement Test originated at AIWS Lumina Lab. Record one milestone for each piece of work. Answer briefly and point to evidence. The digital version may be extended.

1 The work and the starting point

Name · Date · The work · Who benefits · What I could do before beginning

2 Development goal and level of support

Capacity I aim to develop · Roles of AI and of others · Tools and conditions

3 Evidence of advancement

1. **Understanding** — What do I now understand that I did not before?
2. **Judgment** — What did I decide, and why?
3. **Creativity** — What did I contribute?
4. **Command** — How did I lead the work and answer for it?
5. **Elevating the AI Assistant** — Which suggestion fell short, what higher standard did I require, how did the AI respond, and what shows that the assistance improved?
6. **Love** — After this result of working with AI, do I love people and nature more? What insight, feeling, or action shows it?

Tie each statement to evidence or experience; state what has not changed.

4 Value and the Without check

Whom did the result help? What was checked? What did I do with AI, what can I explain on my own, and what can I apply to a new situation?

5 Verdict and next step

Advancement evidenced · Not yet evidenced · Signs of decline — change the practice
One priority capacity · Work to practice it · Date to look back

6 Responsibility and sharing

Who is responsible · Reviewer, if any · What I consent to share, and with whom

Periodic review: Where have I grown? What proves it? What will I develop next?

SOURCES

Beacon Papers

- No. 5, Who May Command? The Double Crisis of Power in the AI Age (2026). bostonglobalforum.org/wp-content/uploads/Beacon_Papers_No5_Who_May_Command.pdf
- No. 6, Who May Author? Human Authorship, AI Assistants, and Responsibility in the Age of Artificial Intelligence (2026). bostonglobalforum.org/news/beacon-papers-no-6-who-may-author/
- No. 7, Human Intelligence Must Lead in the Age of Artificial Intelligence (2026). bostonglobalforum.org/wp-content/uploads/Beacon_Papers_No7_Human_Intelligence_Must_Lead_EN-15-8.pdf
- No. 8, The Architecture of the Civilization of Human–AI Co-Creation (Aug 23, 2026). bostonglobalforum.org/wp-content/uploads/Beacon_Papers_No8_Architecture_of_AIWS_Civilization-official-2.pdf
- No. 10, No AI Agent Without Identity (Sept 6, 2026). bostonglobalforum.org/wp-content/uploads/Beacon_Papers_No10_No_AI_Agent_Without_Identity.pdf
- No. 11, No Power Without a Mandate (Sept 13, 2026). bostonglobalforum.org/wp-content/uploads/Beacon_Papers_No11_No_Power_Without_a_Mandate.pdf
- No. 14, From a Method to a Civilization (Oct 4, 2026). bostonglobalforum.org/wp-content/uploads/Beacon_Papers_No14_From_a_Method_to_a_Civilization.pdf

Boston Global Forum

Boston Global Forum Launches the AIWS Human–AI Co-Creation Movement at the Massachusetts State House (Sept 23, 2026). bostonglobalforum.org/news/boston-global-forum-launches-the-aiws-human-ai-co-creation-movement-at-the-massachusetts-state-house/

Other sources

- G. Liu, B. Christian, T. Dumbalska, M. A. Bakker, R. Dubey, “AI Assistance Reduces Persistence and Hurts Independent Performance,” arXiv:2604.04721 (2026). arxiv.org/abs/2604.04721
- Same study, Section 3.3, “Persistence costs are concentrated among those who obtain direct solutions.” arxiv.org/html/2604.04721v5



AIWS Lumina Lab • Beacon Hill, Boston • Columbus Day, 12 October 2026