



# JMDE AI AND TRANSFORMATIVE EVALUATION IN THE POLYCRISIS

📅 21 Aug 2025

Evaluation takes effort. It has a cost. Generative AI promises to make evaluation radically cheaper and easier to implement. Is this a good thing or a bad thing? In practice, of course, there is no simple answer. Evaluation takes place in many different contexts. It can be *adaptive*, helping people and organisations learn how to develop and optimise public and private activity such as the delivery of services; making evaluative judgements about what (not) to change, how, when, why and for whom. It can involve not only an assessment of facts but also a consideration of values (which? whose?). It can even be *transformative* when xxx. But it can also be *maladaptive* because it uses invalid information or makes invalid judgements or does not reflect the values of stakeholders. And then, there are all the *potential* use cases, where evaluation might take place but does not. These four possibilities are reflected in the rows of the table below. The cells contain some examples:

- A) every day, a teacher informally notices her students' mood and adapts her lesson accordingly. She has been doing it for years and hardly notices she is doing it.
- B1) A bike rental firm asks users to click a smiley or a frowny icon to say how happy they were, but it does nothing with this information.
- B2) As above, but the firm monitors changes in the proportion of frowny icons to identify locations or services which may need intervention. But it uses this to penalise regional staff without a causal understanding of underlying factors.
- C) a city council commissions an evaluation of its recycling programme from an external evaluator.
- D) A VOPE consults its members on how to make national evaluation practice more transformative

	informal	formal / internal	formal / external
transformative		D	
adaptive	A		C
maladaptive		B2	
potential		B1	

What might "introducing AI" mean in these cases? Anything could happen.

In case C the council might decide to use AI both to design and implement a new, realtime system which makes the external evaluator redundant. They might do this by transferring responsibility to an internal

evaluator to manage the process and vouch for its continuing validity. Or a manager might decide to do without any professional input, persuaded by the confident-sounding conclusions coming out of the new AI system. Trust is key. In order to vouch for the robustness of the procedure, what do we rely on?

- an evaluation professional ?
- generic outside advice, e.g. from a management consultancy?
- a software provider (Microsoft? Google?)
- another AI tasked with validating the first?
- nothing at all: we simply rely on the confident-sounding outputs of the AI?

In case B the firm might redesign its system to use AI to make better, automated, realtime judgements about issues, causes and remedies in a way which is transparent and leaves room for employees and staff to respond, discuss and negotiate. Here we can dimly see a possible new role for evaluation professionals, to help co-design (and vouch for) new automated or semi-automated evaluation systems.

In case A it is not hard to imagine an AI system which monitors the students' social media feeds, tone of voice, eye movements as well as performance on set tasks, and gives the teacher realtime suggestions for pacing and adapting the class overall and for individuals. But most parents, students and teachers would (at the moment) likely be horrified at such a suggestion and it is doubtful if could be adaptive.

The bottom line in our table contains the largest potential (for adaptation, for transformation but also for poor or counterproductive solutions). At the extreme, a manager might engage an AI agent to sift through an entire organisation's workflows, identify areas which are currently not evaluated at all, and for each one, engage AI agents to suggest interoperable, cheap and robust solutions, and then connect them to tools to actually implement this new comprehensive suite of evaluation services. Of course, the implementation on paper might be laughably cheap, the real-life implementation is all about the human interface and might be arbitrarily disruptive and expensive and maladaptive.

The day is probably not far off when a manager can open their computer one morning to be greeted with a message like (n.d.) morning, I have created a system-level valuation suite while you were sleeping. Would you like to switch it on or do you want to bother with reviewing it first? Assuming you would like to switch it on, will you want it to work completely independently or do you want to be informed about top-level decisions like hiring and firing?