European Funds

Use of GenAI in the design of the Evaluation of the implementation of the partnership principle in the 2021-2027 perspective (case study from Poland)

Marlena Stępień, Ministry of Development Funds and Regional Policy, Poland

European Funds

Republic of Poland

Co-funded by the European Union

Key assumptions of the design process

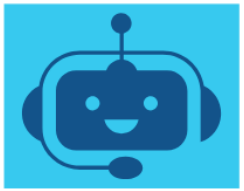
COOPERATION WITH PARTNERS

Invite representatives of partner organizations (through an open call) to cooperate throughout the entire evaluation process

Result: two representatives engaged into the entire evaluation process

Involvement of stakeholders (including partners) from the verybeginning – "white sheet method"

Result: over 500 proposed research issues and questions



USE OF NEW TECHNOLOGIES

Designing the evaluation study using GenAI

Prior

Collection of proposals for questions and issues from stakeholder (500+ submissions)

PROCESS FLOW

How was the evaluation study designed using GenAI?

Preparation& intake

ChatGPT

identifying sources of knowledge, initial data cleaning

First-pass structuring

NotebookLM

grouped all the questions into 13 thematic categories

Iterative synthesis

ChatGPT

iterative work on research questions until the final 20 were formulated

Creation of the Terms of Reference

ChatGPT

Stakeholders validation

central administration, local government units, NGOs, social and economic partners

Final ToR

3 working days of the evaluation coordinator (from NEU)

One-month stakeholders consultations (first round of comments -> amendments -> second round of comments)

HUMAN-IN-THE-LOOP

Supervision of the whole process by the evaluation coordinator (from NEU)

What GenAI did well

(with tangible gains):

RAPID SYNTHESIS AT SCALE

Turned 500+ stakeholders inputs into a coherent, prioritized set of evaluation questions and a draft ToR in ~3 working days (work that typically takes several weeks).

STRUCTURED THINKING

Delivered fast clustering, deduplication, and gap-spotting, helping us move from “white noise” to thematic clarity; improved traceability by saving iterations (prompts & outputs).

SIMPLIFIED DESIGN

Produced variants of methods, criteria, data sources, and schedules, which sped up deliberation and made choices explicit.

STAKEHOLDER RESONANCE

The use of GenAI allowed all stakeholders proposals to be taken into account (without omission).The AI-supported ToR was perceived as “fresh” and “interesting”.

RESOURCE EFFICIENCY

Without the use of GenAI, it would have been necessary not only to prolong the process of creating the ToR, but also to involve many more people in the work than a single coordinator.

Where GenAI struggled

(and how the evaluation coordinator mitigated):

OVER-COMPRESSION AND OMISSIONS IN EARLY, "ALL-AT-ONCE" ANALYSIS	Mitigation: switch to chunked processing (one theme at a time) and document-based work (NotebookLM first, then ChatGPT).
INCONSISTENT GRANULARITY ACROSS THEMES AND OCCASIONAL OVERCONFIDENT GENERALISATIONS	Mitigation: coordinator harmonised levels of detail and verified claims against source documents.
PROMPT SENSITIVITY. OUTPUT QUALITY DEPENDED HEAVILY ON HOW QUESTIONS WERE FRAMED	Mitigation: iterative prompt refinement and explicit instructions on scope and constraints.
RISK OF BIAS / CROWDING OUT NICHE TOPICS	Mitigation: manual re-insertion of under-represented issues during reviews; transparent decision logs.

On working with Wizzards

"This is the issue with wizards: We're getting something magical, but we're also becoming the audience rather than the magician, or even the magician's assistant."

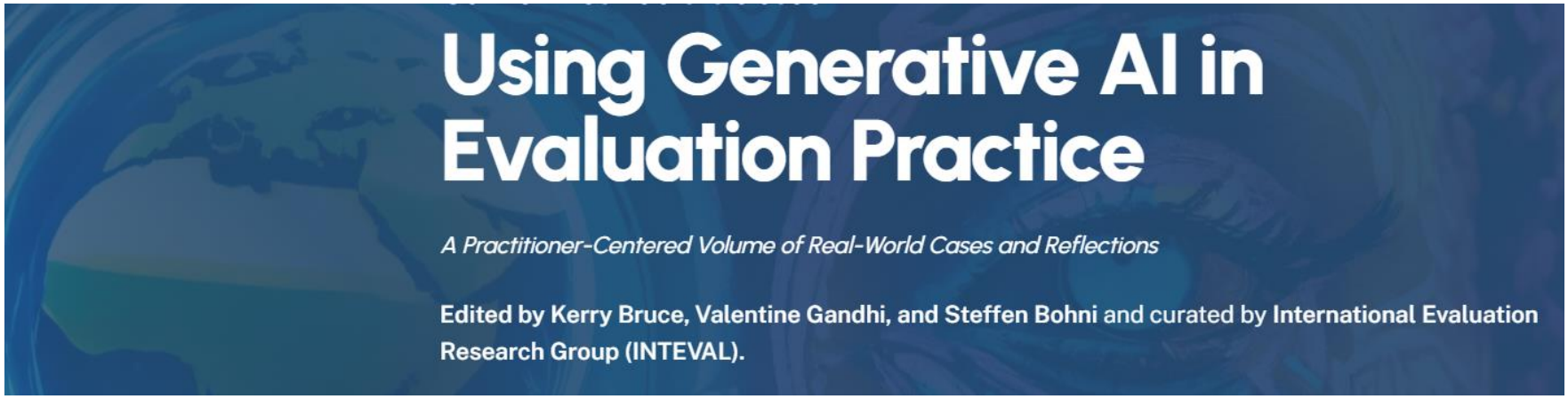
On Working with Wizards. Verifying magic on the jagged frontier – Ethan Mollick, 2025, oneusefulthing.org

On working with Wizzards

"In the co-intelligence model, we guided, corrected and collaborated. Increasingly, we prompt, wait, and verify… if we can."

On Working with Wizards. Verifying magic on the jagged frontier – Ethan Mollick, 2025, oneusefulthing.org

New open-access book on using GenAI in evaluation practice – coming in early 2026



https://www.dev-cafe.org/book-ai4mel/

Scoping evaluation with GenAI: A case study of the partnership principle in Poland (Marlena Stępień)

18.11.2025

European Funds

Thank you!

marlena.stepien@mfipr.gov.pl

ewaluacja@mfipr.gov.pl (National Evaluation Unit)

Linkedin.com/in/marlenastepien

European Funds

Republic of Poland

Co-funded by the European Union