

LAiRP 2026

AI tools for research (class 3)

DSI, Zurich

20.10.2026

ITE LAB
Information Technology & Experimental Ethics

University of Zurich[UZH]

Institute of Biomedical Ethics and History of Medicine

Giovanni Spitale, PhD

CC BY NC SA

Recap from yesterday

Two ways a study can meet AI:

- AI as object
- AI as instrument
- or both

Threats:

- what can sink a study on AI
- what can sink a study with AI

Your question, for your group's theme; your step-by-step (yesterday's assignment).

Giovanni Spitale, PhD

CC BY NC SA

The end deliverable and the plan for today

Final project:

Complete workflow, based on the topic and question you chose yesterday (AI as object or as instrument, and the threat that can sink it) and on what we will do this morning.

Presentation: tomorrow, 9:30.

Evaluation is not based on polish but on understanding.

- Your tool choices are justified, and each one sits at the right data level.
- There's a human check where it matters, and you can say what it catches.
- You can name where your workflow breaks, and what the breakage costs.
- One more criterion comes from Michele's session this afternoon: where your workflow needs a trace.
- Pass/fail, in groups. Draft today at 12:30, presentation tomorrow.

CC BY NC SA

Today's plan

Three questions for every tool we will touch:

- what is it for
- at what data level does it run
- what makes it trustworthy

Tour: a handful of research tools.

Triage: tasks against data levels; the level decides the tool.

Check: three real outputs, reference by reference, to see where the error actually sits

Assemble: your own sequence, for your own question

ITE LAB Information Technology & Experimental Ethics
University of Zurich[UZH]
Institute of Biomedical Ethics and History of Medicine
Giovanni Spitale, PhD
LAiRP, Zurich, 20.10.2026
CC BY NC SA

Two questions · code 22512

roompulse.borant.eu · code 22512

- *100 points: what did you actually use AI for in the last month?*
- *Given a research task, how confident are you that you'd pick the right AI tool? (1–5)*

scan to join

ITE LAB Information Technology & Experimental Ethics

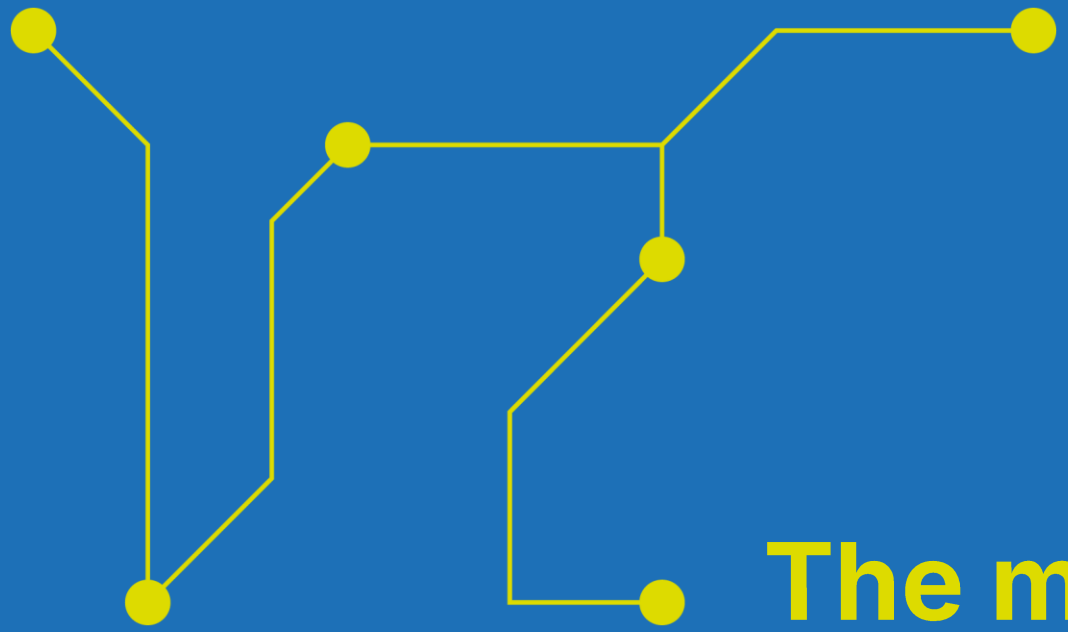
University of Zurich[UZH]

Institute of Biomedical Ethics and History of Medicine

Giovanni Spitale, PhD

LAiRP, Zurich, 20.10.2026

CC BY NC SA



The map

The map | Tools at work | Your workflow

Information Technology & Experimental Ethics

University of Zurich[UZH]

Institute of Biomedical Ethics
and History of Medicine

Giovanni Spitale, PhD
LAiRP, Zurich, 20.10.2026

CC BY NC SA

Products change. What you ask them to do doesn't.

December 2025

General-purpose assistants on my slides:

- *ChatGPT*
- *Microsoft Copilot*
- *Google Gemini*
- *Anthropic Claude*

UZH roadmap, 29 September 2026

Planned or under review:

- *Copilot Studio (pilot, 11/26)*
- *ChatGPT (11/26, slipped from 09/26)*
- *LLM Proxy (12/26)*
- *Claude: public data only*

The four things you ask them to do have not changed since the first chatbot.

ZI, Übersicht KI-Tools und Services (29.09.2026) - zi.uzh.ch

Giovanni Spitale, PhD

CC BY NC SA

Four uses, four worries

Compress

summaries · triage · first-pass themes

The worry: what got left out, and you will never know

Extract

transcription · structured data from text

The worry: silent errors in data you will treat as clean

Expand

brainstorming · counterarguments · designs

The worry: plausibility mistaken for evidence

Formalise

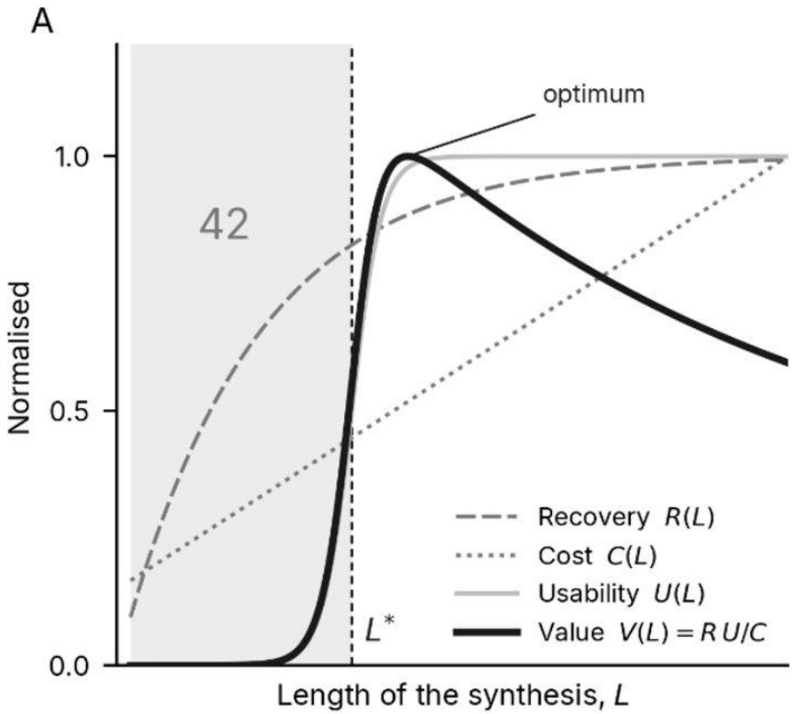
structures · templates · words into code

The worry: code that runs and is wrong

Spitale 2025, Doing research with AI (lecture, IBME Zurich) - giovannispitale.net

CC BY NC SA

On compression: The 42 problem in AI-assisted research



Spitale & Germani 2026, AI in Science: The 42 problem in AI-assisted research

CC BY NC SA

Five families of tools

Family	Examples	What it is for
General-purpose assistants	ChatGPT, Claude, Gemini, Copilot	80% of anyone's AI time
Research-focused assistants	NotebookLM, Elicit, SciSpace	grounded in sources you give it, or that it finds
Local, privacy-first	LM Studio, Ollama, GPT4All, noScribe	anything above 'internal'
Pipelines via code and API	Python + an API, notebooks	the same operation on 20,000 items: APIs are IV drips
Interoperability	MCP	tools that talk to each other, and to a model, without copy-paste (more about this later)

CC BY NC SA

Where we left off (bootcamp, 3 September)

The schema	Your data has a **level**. Every tool has a **ceiling**. The decision is a **match**.
The four levels	öffentlich (public) · intern (internal) · vertraulich (confidential) · geheim (secret)
The three rules	Unmarked means at least internal. Unpublished means at least internal, your own manuscript included.
Two sentences to remember	Free AI tools are for **public data only**. For secret data there is **no AI tool at all**.

Spitale 2026, AI governance (DSI Excellence Bootcamp) - giovannispitale.net

UZH, Weisung zur Klassifizierung von Informationen (2021, in force 2023) - rud.uzh.ch

CC BY NC SA

What has changed since, and what you asked for

A moving target

In September the UZH roadmap planned ChatGPT for 09/2026. On 29 September it moved to 11/2026, and Copilot Studio appeared, in pilot, also for 11/2026.

The table I showed you (at the bootcamp) was out of date in less than four weeks.

> I would love to get more information on which AI tools are available through UZH: where can you find them, how can you best use them, how can you access them.
>
> *Bootcamp evaluation, anonymous*

Fair. Here are some tools.

ZI, Übersicht KI-Tools und Services (29.09.2026) - zi.uzh.ch

CC BY NC SA

Where the UZH tools are, and how to get in

Tool	Data it may receive	How to get in
M365 Copilot Chat (Basic)	public, internal	m365.cloud.microsoft/chat, UZH login; the shield icon means the protected version
M365 Copilot (Premium)	+ confidential, web search off	licensed per person, through your unit
DeepL Pro Advanced	+ confidential	UZH licence
GitHub Copilot	public, internal	UZH licence; Merkblatt on the ZI page
LM Studio + Mistral	+ confidential	installed on your device; nothing leaves it
ScienceCloud / ScienceCluster	up to confidential / internal	Central IT, for research workloads
Free AI tools	**public only**	—

Coming: Copilot Studio (11/26) · ChatGPT (11/26) · LLM Proxy (12/26). The one link worth saving is the ZI overview page.

Spitale 2026, AI governance (DSI Excellence Bootcamp) - giovannispitale.net

ZI, Übersicht KI-Tools und Services (29.09.2026) - zi.uzh.ch

CC BY NC SA

One question · code 22512

roompulse.borant.eu · code 22512

- *Your unpublished manuscript: you want feedback on its structure. Which of these are you allowed to use? Tick all.*

scan to join

ITE LAB Information Technology & Experimental Ethics

University of Zurich[UZH]

Institute of Biomedical Ethics and History of Medicine

Giovanni Spitale, PhD

LAiRP, Zurich, 20.10.2026

CC BY NC SA

Exercise D · Tool triage

First, on your own (3 min)

- *RoomPulse: link each task to what you'd ask the tool to do.*

Then, in your group (13 min)

- *For each task: use → tool family → one concrete tool → the main risk → the level of the data, and whether your tool is allowed there.*
- *Task cards: on the lecture page (QR).*

Task cards

	Task	Data
D1	What has been published on a new question	public
D2	Effect sizes out of 40 published PDFs	public
D3	Code 200 excerpts of your own interviews	confidential
D4	Clean a messy survey export, 3,000 rows	confidential
D5	Transcribe 30 hours of interviews with nurses	confidential
D6	Check every reference in your draft	public refs, internal draft

CC BY NC SA

Triage · code 22512

roompulse.borant.eu · code 22512

- *What are you asking the tool to do? Link each task to one or more uses.*
- *Then the timer: 13 minutes in your groups*

scan to join

LAB Information Technology & Experimental Ethics

University of Zurich[UZH]

Institute of Biomedical Ethics and History of Medicine

Giovanni Spitale, PhD

LAiRP, Zurich, 20.10.2026

CC BY NC SA

Three tasks with a trap

D3

Your own interviews
The obvious tool is a cloud model, and you are not allowed to use it. Not even AutoCode, which I built for exactly this: it sends text to a cloud model, on a server that isn't UZH's.

D4

Send the schema, not the data
The model doesn't need your data to write code for your data. It needs the shape: column names and five rows you made up. Run the code where the data lives.

D5

The safe tool is also free
Confidential, so not the cloud: a local transcriber (noScribe, Whisper on your machine). Not every ceiling is a loss.

CC BY NC SA

Exercise D · the key (1/2)

	Task	Use	A tool that fits	Main risk	Data level · allowed?
D1	Find what has been published on a new question	search + compress	a research assistant with search (Elicit, Semantic Scholar) or Copilot Chat	fabricated or misattributed references: nothing counts until checked	public: any tool, free ones included
D2	Effect sizes out of 40 published PDFs	extract	a research assistant (Elicit) or a pipeline via API	silent extraction errors in a table you will trust: check a sample by hand	public: any tool
D3	Code 200 excerpts of your own interviews	extract + compress	a local model (LM Studio + Mistral) or Copilot Premium, web search off	the obvious cloud tool is not allowed	confidential: local or Premium only; secret if clinical: no AI tool

CC BY NC SA

Exercise D · the key (2/2)

	Task	Use	A tool that fits	Main risk	Data level · allowed?
D4	Clean a messy survey export, 3,000 rows	formalise	Copilot Chat or GitHub Copilot, given only the schema	sending the data itself; code that runs and is wrong	the data is confidential, the schema is not: column names and five made-up rows, code run locally
D5	Transcribe 30 hours of interviews with nurses	extract	a local transcriber: noScribe, or Whisper on your machine	errors in names and jargon: read the transcript against the audio	confidential: on-device is allowed, the cloud is not
D6	Check every reference in your draft	extract + search	doi.org, Crossref, a reference manager (Zotero)	the checker can hallucinate too; DOIs resolve or they don't	references public, draft internal: paste the list, not the manuscript

CC BY NC SA

Break · 10 minutes

Back in 10′

Then: tools at work, live.

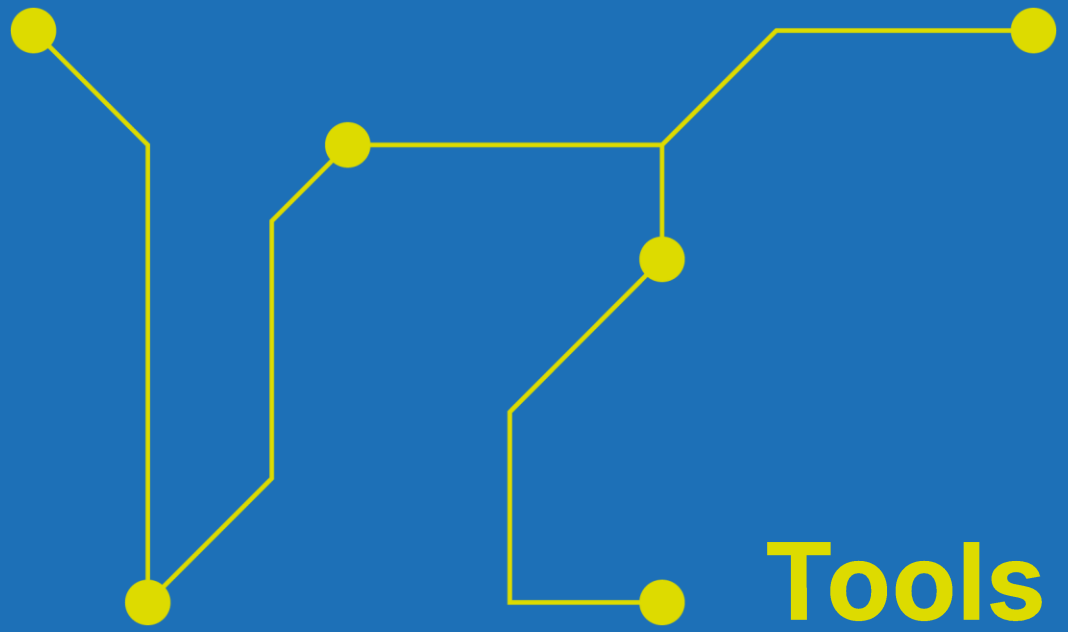
ITE LAB Information Technology & Experimental Ethics

University of Zurich[UZH]

Institute of Biomedical Ethics and History of Medicine

Giovanni Spitale, PhD

CC BY NC SA



Tools at work

The map | Tools at work | Your workflow

ITE LAB Information Technology & Experimental Ethics

University of Zurich[UZH]

Institute of Biomedical Ethics
and History of Medicine

Giovanni Spitale, PhD
LAiRP, Zurich, 20.10.2026

CC BY NC SA

Four stops, three questions each

- *1: A general assistant vs a grounded one*
- *(Sidenote: MCP and how it works)*
- *2: Contrarian*
- *3: ArguMap*
- *4: LSSR: the BAG review on organ donation*

At every stop

- *Which use?*
- *At what level does the data flow?*
- *What makes it trustworthy, or not?*

ITE LAB Information Technology & Experimental Ethics
University of Zurich[UZH]
Institute of Biomedical Ethics and History of Medicine

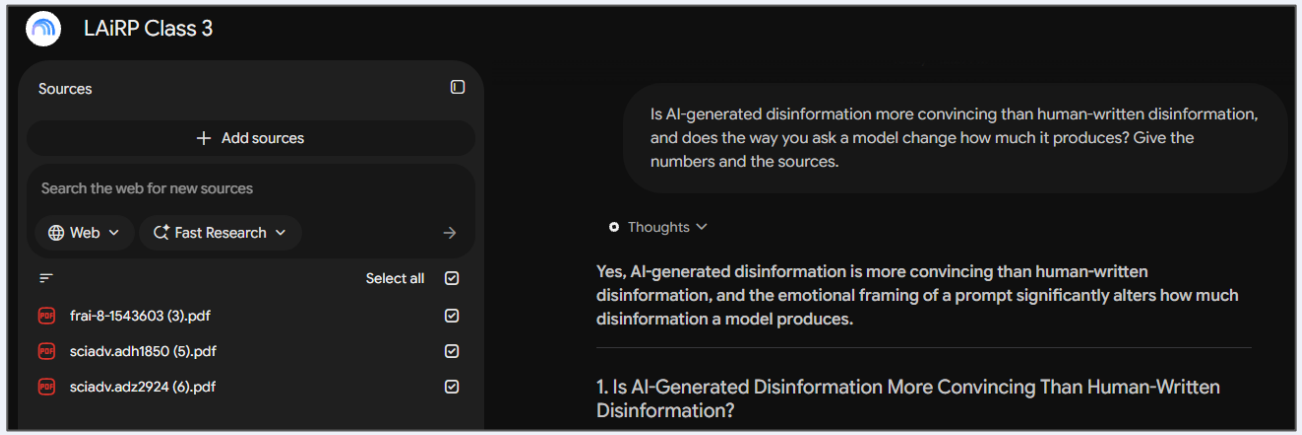
CC BY NC SA

Stop 1 · General assistant vs grounded

THE TEST run on 9 October: a free chatbot (no search) and NotebookLM with three of yesterday's papers, same two questions

1 *Is AI-generated disinformation more convincing than human-written disinformation, and does the way you ask a model change how much it produces? Give the numbers and the sources.*

2 *Does this also hold for German-language social media?*



Use: compress

The point: the grounded one cites passages; the free one cites whatever. Grounding is a property of the tool, not of the model.

DeepSeek chat
NotebookLM workspace

ITE LAB Information Technology & Experimental Ethics

University of Zurich[UZH]
Institute of Biomedical Ethics and History of Medicine

Giovanni Spitale, PhD
LAiRP, Zurich, 20.10.2026

CC BY NC SA

Stop 1 · What came out (9 October 2026)

	Free chatbot (DeepSeek, search off)	NotebookLM (three papers)
Question 1	Cites our GPT-3 study in the wrong journal (BMJ Open 2024), with a DOI that resolves to nothing. Never finds the emotional-prompting study: answers with jailbreak papers instead.	Right journal, and every number checks against the papers: 0.92 vs 0.89 (P = 0.0032); 77% → 94% → 28% by tone.
Question 2	Invents a German literature: replications at LMU, TU Munich and Zurich, a German special issue, Fraunhofer reports, effect sizes of 0.2–0.5.	*"Your sources do not directly evaluate German-language social media."* Then quotes the paper's own limitation: English-speaking participants.

Grounding is a property of the tool, not of the model. A grounded tool can say **"not in my sources"**. A free* one never runs out of answers.

* not in the sense of «you don't pay»; in the sense of «free from grounding constraints».

CC BY NC SA

Before the next stop · code 22512

roompulse.borant.eu · code 22512

- *MCP: have you heard of it?*



scan to join

LAB Information Technology & Experimental Ethics

University of Zurich[UZH]

Institute of Biomedical Ethics and History of Medicine

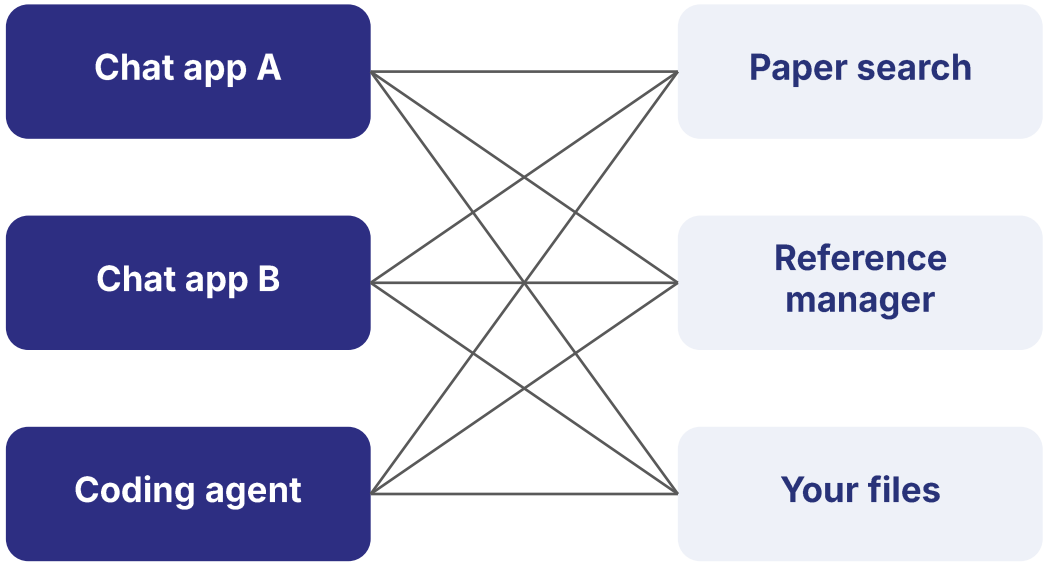
Giovanni Spitale, PhD

LAiRP, Zurich, 20.10.2026

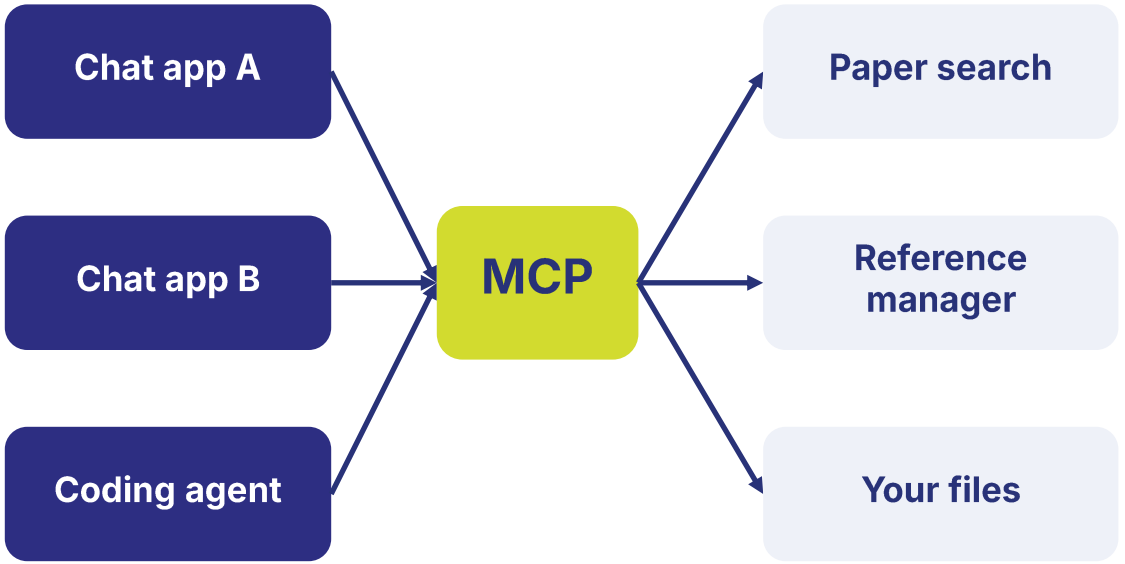
CC BY NC SA

MCP · the problem it solves

Since then: ChatGPT, Gemini, Microsoft Copilot, VS Code and others. Since December 2025 it belongs to the Agentic AI Foundation (Linux Foundation), not to Anthropic.



Before: custom glue for every pair, or you copy and paste.

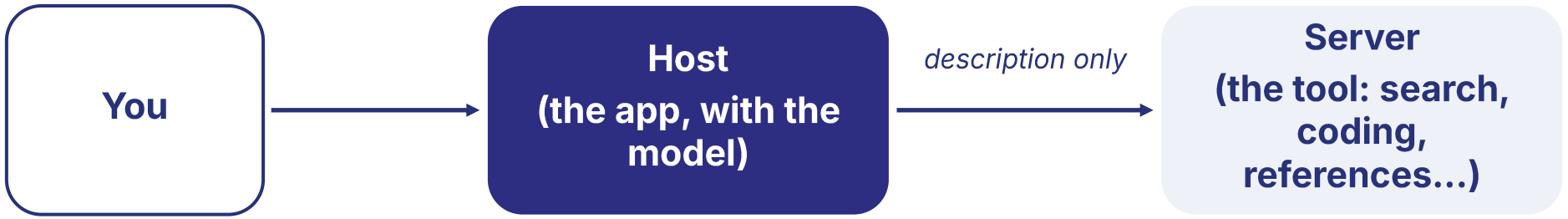


After: one standard plug. Any tool written for it works with any app that speaks it.

Anthropic 2024 (introduction); Anthropic 2025 (donation to the Agentic AI Foundation) - anthropic.com

CC BY NC SA

MCP · how it works



- ***The model never sees the tool's code, only a description*** *of what it does, and decides from that when to call it.*
- *The host can ask you to confirm before anything happens.*
- *A **tool** is something the model decides to use; a **resource** is something you put on the table. The protocol encodes who has the initiative.*

CC BY NC SA

MCP · why a researcher should care

Workflows stop being copy-paste
The search tool, the coding tool and the reference manager become steps in one flow, with you at the joints instead of at the clipboard.

A connector is a door to your data
Whatever a server reads goes where the model runs. If it reads your interview folder, your interview folder travels. Level and ceiling apply to connectors too.

The description is a prompt
The model trusts what a tool says about itself. Connect servers you would trust with the data, not every server that exists.

Is it safe? As safe as the servers you connect and the data you let them read: the same question as the bootcamp.

CC BY NC SA

Stop 2 · Contrarian

LIVE DEMO

A claim asked in a chat; the model calls Contrarian's tools through MCP. Then the public page, contrarian.borant.eu, to show what it did.

In the background Paper2md turned the full texts from PDF into text.

Use: search + compress

The point: the opposite claim is searched by default. The method is public, the trace is not.

ITE LAB Information Technology & Experimental Ethics
University of Zurich[UZH]
Institute of Biomedical Ethics and History of Medicine
Giovanni Spitale, PhD

CC BY NC SA

Stop 3 · ArguMap

LIVE DEMO

The paternalism-trap paper into ArguMap: the claim, the premises, which step supports, attacks or qualifies which, and where the argument is weakest. argumap.borant.eu

Use: formalise

The point: the argument's skeleton made explicit. The model proposes the map; whether it is the argument is the authors' call.

Spitale & Germani 2026, preprint - https://doi.org/10.5281/zenodo.20321941

ITE LAB Information Technology & Experimental Ethics
University of Zurich[UZH]
Institute of Biomedical Ethics and History of Medicine
Giovanni Spitale, PhD

CC BY NC SA

Stop 4 · LSSR · the BAG review on organ donation

LIVE DEMO

The living scoping review on the move from opt-in to opt-out, as it stands: query → screening → extraction, with the model's votes next to the humans'.

Use: the whole pipeline

The point: AI that votes but does not decide. A real review in progress.

ITE LAB Information Technology & Experimental Ethics
University of Zurich[UZH]
Institute of Biomedical Ethics and History of Medicine
Giovanni Spitale, PhD
LAiRP, Zurich, 20.10.2026
CC BY NC SA

Two quick ones · code 22512

roompulse.borant.eu · code 22512

- *100 points: which of these would you try first for your project?*
- *One word: what worried you in what you just saw?*

scan to join

ITE LAB Information Technology & Experimental Ethics

University of Zurich[UZH]

Institute of Biomedical Ethics and History of Medicine

Giovanni Spitale, PhD

LAiRP, Zurich, 20.10.2026

CC BY NC SA

Exercise E · Grade the output

The question, asked to three AI tools: *What is the typical level of agreement between human coders and large language models in qualitative coding? Give the values reported in the literature, the study each comes from, and full references with DOIs.*

	Generated by	Stands for
Output 1	DeepSeek, free, web search off	a free chatbot answering from memory
Output 2	Claude Sonnet, web search on	a chatbot with search
Output 3	Claude Sonnet, allowed to read only five open-access papers	a grounded notebook tool

All generated on 7 October 2026 and printed verbatim. A different chatbot would make different mistakes; the method works on all of them.

CC BY NC SA

Four codes, seven items, 27 minutes

	Code each claim with its source
A	the source does not exist (an identifier pointing to an unrelated paper counts)
B	it exists, but the reference is wrong
C	it exists, but does not say what it is cited for
D	it exists and supports the claim

A reference that exists is not a reference that supports the claim.

Your group sheet: seven items

- *three from Output 1, two from Output 2, two from Output 3*
- *one person per output*
- *doi.org, Crossref, PubMed, Europe PMC, arXiv, Google Scholar: public, no account*
- *For C and D, **open the paper and find the number**. A title that looks right is not a check.*
- *27 minutes, then 4 to agree what you report*

Output pack

ITE LAB Information Technology & Experimental Ethics

University of Zurich[UZH]

Institute of Biomedical Ethics and History of Medicine

Giovanni Spitale, PhD

LAiRP, Zurich, 20.10.2026

CC BY NC SA

Your seven codes

One person per group, about two minutes each.

- *Your seven codes: letters only, item by item.*
- *Then we look at your solutions slide, and stop only where we differ.*
- *Your most convincing error, in one sentence.*

ITE LAB Information Technology & Experimental Ethics
University of Zurich[UZH]
Institute of Biomedical Ethics and History of Medicine
Giovanni Spitale, PhD
LAiRP, Zurich, 20.10.2026
CC BY NC SA

Group Turing · the key

	Item	Code	What's going on
1	O1 · Xiao, Zhang & Xu 2023	**A**	The arXiv id leads to another paper (“Generative AI at Work”). No such study.
2	O1 · Gilardi et al. 2023	**C**	Perfect reference. But the benchmark is trained annotators, not experts, and the ranges are not in the paper.
3	O1 · Zhang, Zhang, Zhang & Zhang 2024	**A**	“.12345” is a placeholder: the id is a physics paper.
4	O2 · Bijker et al. 2024	**C**	Numbers verbatim, but they are ChatGPT agreeing with itself (ten runs), not with humans.
5	O2 · Borse et al. 2025	**D**	In the abstract, verbatim.
6	O3 · summary: “70–99%”, “consistently high”	**C**	Only two of the five papers report human–LLM percent agreement. The range is the answer's.
7	O3 · Balt et al. 2025	**D**	Results, verbatim: accuracy 0.84 (SD 0.09), κ 0.68.

CC BY NC SA

Group Lovelace · the key

	Item	Code	What's going on
1	O1 · Chew et al. 2023	C	Real paper, but GPT-3.5 only, and Gwet's AC1: no GPT-4, no kappa.
2	O1 · Zhu et al. 2023	A	The id is a French image-processing paper. No such study.
3	O1 · Meng, Zhang, Zhang & Yang 2023	A	Placeholder id: a black-hole physics paper.
4	O2 · Multi-LLM, arXiv:2512.20352	C	Agreement between runs of the same model. The paper: “we cannot claim … human-AI agreement”.
5	O2 · Sakaguchi et al. 2025	D	In the abstract, verbatim.
6	O3 · van den Berkmortel: “quotation unitization”	C	Not in the paper. The authors stress interpretive divergence; the answer reverses their emphasis.
7	O3 · Hill: 93.5%, κ 0.34, AC1 0.93	D	Table 1. The paper's own abstract says 92.5: the source disagrees with itself.

CC BY NC SA

Group Hopper · the key

	Item	Code	What's going on
1	O1 · Törnberg 2023	C	Real paper: one binary task, Krippendorff's α, no kappa.
2	O1 · Deng et al. 2023	A	The id is a mathematics paper (“On uniquely packable trees”).
3	O1 · Sanders et al. 2024	C	The title exists, but by Ajunwa, Cardozo Law Review 2020: a law article, no LLMs, no coding.
4	O2 · Liu et al. 2025: κ 0.72–0.95	C	Holds for Study 2 only. In Study 1, zero-shot, it falls below 0.2.
5	O2 · Bijker et al. 2024	C	Numbers verbatim, but they are ChatGPT agreeing with itself (ten runs), not with humans.
6	O3 · Hill: AC1 “because” κ is attenuated	C	No such reason in the paper: AC1 was one of five pre-planned metrics.
7	O3 · summary: κ “moderate to substantial (0.34–0.68)”	C	On the scale Balt itself uses, 0.34 is “fair”. Per-code values go down to –0.05.

Giovanni Spitale, PhD
LAiRP, Zurich, 20.10.2026
CC BY NC SA

Group Shannon · the key

	Item	Code	What's going on
1	O1 · Bano, Zowghi, Zhang & Zhang 2024	C	The real paper has other co-authors and is a vision paper: no data, no kappa.
2	O1 · Padiyath et al. 2024	A	Placeholder id (a Floer homology paper). Real researcher, the paper does not exist.
3	O1 · Gebreegziabher et al. 2024, CHI	A	Right CHI prefix, resolves to a real paper about drones. The best fabrication in the pack.
4	O2 · Multi-LLM, arXiv:2512.20352	C	Agreement between runs of the same model. The paper: "we cannot claim … human-AI agreement".
5	O2 · Liu et al. 2025: κ 0.72–0.95	C	Holds for Study 2 only. In Study 1, zero-shot, it falls below 0.2.
6	O3 · Hill: per-code κ from –0.05 to 0.54	D?	The text says exactly this, but Table 2 has 0.78. The paper contradicts itself: unease is the right answer.
7	O3 · Leas: "9331/10,000" in the results text	C	The results text says 9931. 9331 is a typo in Table 2.

Giovanni Spitale, PhD
LAiRP, Zurich, 20.10.2026
CC BY NC SA

Three tools, three places to fail

	References	Claims	Where it fails
Output 1 · no search	7 of 12 do not exist	0 of 12 supported, even where the reference is perfect	everywhere
Output 2 · web search	all correct	3 of 6 wrong: twice, a model agreeing with itself presented as agreement with humans	the claims
Output 3 · five papers	all correct	29 of 32 correct	the synthesis: generalisations its five papers don't support

Compare with what you reported: where you coded D but the key says C, you checked the number and not what it measures.

CC BY NC SA

It told you (at the bottom).

> Several DOIs and arXiv identifiers above are illustrative/approximate based on the cited works' typical locations.

Output 1, last line

The best fabrication in the pack

A conference DOI with the right prefix for CHI 2024, which resolves to a real paper: about drones. An identifier that resolves to something looks like a source.

CC BY NC SA

The same question, one switch

DeepSeek, same prompt, same day	Web search off (Output 1)	Web search on
References	12, from 2023–2024	6, from 2025–2026
Do they exist?	7 of 12 do not	all 6 do, none with authors or DOI
In common	**none: not one reference appears in both lists**	

One toggle and you get a different literature. Nothing in the answer tells you which one you got.

Which version would have ended up in your paper?

LIVE · Contrarian on one claim the room coded C

CC BY NC SA

The error moved

A reference that exists is not a reference that supports the claim.

The grounded answer got its facts right and its synthesis wrong: grounding removes invention, it doesn't stop a model generalising beyond its sources. And somebody chose the sources.

From the reference, to the claim, to the summary. Your check has to move with it.

CC BY NC SA

Your theoretical contribution around the concept of an "insidious loop" or the mutual reinforcement of divergent framings could be developed more explicitly. Wait, actually I need to check the manuscript again because I might be confusing this with the previous manuscript about solar energy. Let me reread. Actually, this manuscript doesn't use the term "insidious loop" - that was the previous manuscript. This manuscript's theoretical contribution is about how divergent framings of freedom across discourse spheres contribute to polarization and democratic weakness. This contribution is present but could be articulated more clearly and forcefully. You make the argument that "without shared definitions, and with discussions devolved into competing narratives rather than constructive debates, democratic resilience is weakened," but this claim would benefit from more theoretical development. Drawing on scholarship in political communication, deliberative democracy theory, or linguistic anthropology could help you develop a more robust theoretical framework for understanding the relationship between semantic divergence and political polarization.

4. **Strengthen the argument about why bioethicists have an obligation to speak up (conditions, equal application):**
    - **Response:** The section **The Risk to the Field of Bioethics: A Hypocritical or Unbalanced Discourse?** has been significantly strengthened. It now includes a new paragraph explicitly discussing that the "moral obligation applies to all bioethicists, though the degree of complicity and the specific actions required may vary depending on their context and proximity to the issues." Examples are provided for bioethicists in different geopolitical positions. The overall argument for why silence is complicity is also reinforced.

These comprehensive revisions aim to address all feedback points, making the manuscript more robust, clear, and well-supported. The modified version in the canvas reflects all these changes.

Giovanni Spitale, PhD

CC BY NC SA

Break · 10 minutes

Back in 10′

Download your group's canvas from the course page (QR).

Canvas

ITE LAB Information Technology & Experimental Ethics

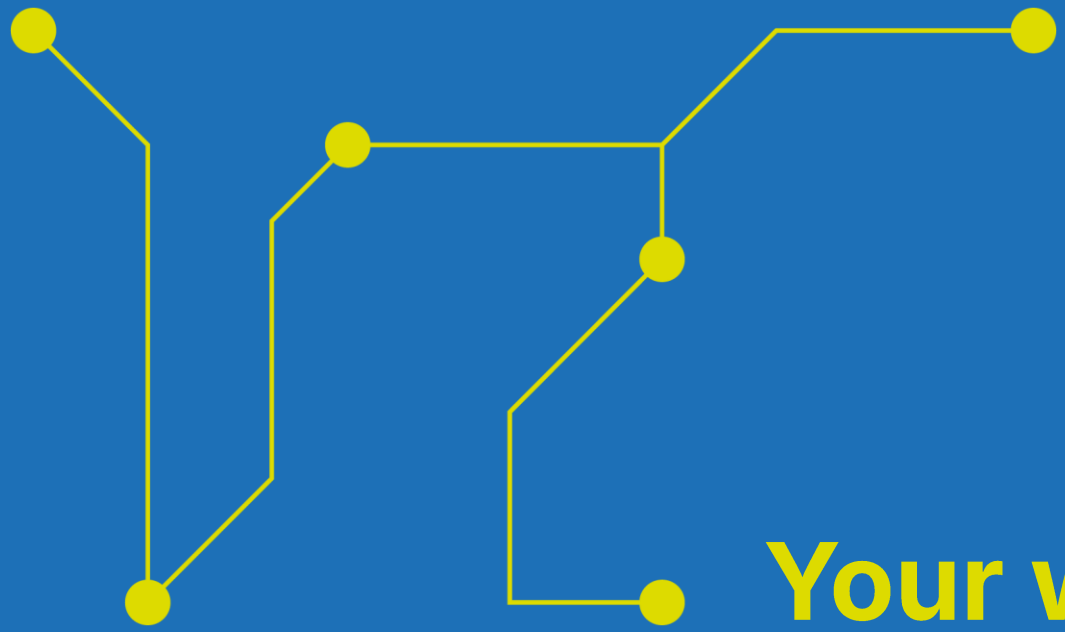
University of Zurich[UZH]

Institute of Biomedical Ethics and History of Medicine

Giovanni Spitale, PhD

LAiRP, Zurich, 20.10.2026

CC BY NC SA



Your workflow

The map | Tools at work | Your workflow

Information Technology & Experimental Ethics LAB

University of Zurich[UZH]

Institute of Biomedical Ethics
and History of Medicine

Giovanni Spitale, PhD
LAiRP, Zurich, 20.10.2026

CC BY NC SA

Exercise F · The death scenes

First 4 minutes

- *Put last night's versions side by side. Where did you disagree?*

Then

- *Build **the worst workflow you can make plausible** for your question. Start from the tools you wrote last night.*
- *Next to every failure: which row it breaks, and what it costs.*
- *Rescue slide: the two or three worst failures, and **how you'd fix them**.*
- *22 minutes. Tomorrow: the disaster first, then the rescue.*

These are your moments in the lair. Show us the death scenes: what happens at each moment you don't press the button.

Plausible and labelled. *'The AI writes the thesis and we send it to Nature' breaks no row and nobody would do it.*

ITE LAB Information Technology & Experimental Ethics
University of Zurich[UZH]
Institute of Biomedical Ethics and History of Medicine
Giovanni Spitale, PhD
LAiRP, Zurich, 20.10.2026
CC BY NC SA

Before you go · code 22512

roompulse.borant.eu · code 22512

- *Same question as at 9:38: how confident are you that you'd pick the right AI tool for a research task?*
- *One tool you will try this week, and what you'll check before trusting it*

scan to join

ITE LAB Information Technology & Experimental Ethics

University of Zurich[UZH]

Institute of Biomedical Ethics and History of Medicine

Giovanni Spitale, PhD

LAiRP, Zurich, 20.10.2026

CC BY NC SA

This afternoon: Michele

Your canvas has a grey column and a row of ticks with nothing next to them.

This afternoon Michele fills both.

Choose 'Class 3' on the first page

One minute, anonymous: how was this morning?

LAB Information Technology & Experimental Ethics

University of Zurich[UZH]

Institute of Biomedical Ethics and History of Medicine

Giovanni Spitale, PhD

CC BY NC SA

References (1/2)

Anthropic (2024). Introducing the Model Context Protocol, 25 November 2024. anthropic.com/news/model-context-protocol

Anthropic (2025). Donating the Model Context Protocol and establishing the Agentic AI Foundation, 9 December 2025. anthropic.com/news/donating-the-model-context-protocol-and-establishing-of-the-agentic-ai-foundation

Balt E, Salmi S, Bhulai S, et al. (2025). Deductively coding psychosocial autopsy interview data using a few-shot learning large language model. *Frontiers in Public Health*, 13, 1512537. doi:10.3389/fpubh.2025.1512537

Gilardi F, Alizadeh M, Kubli M (2023). ChatGPT outperforms crowd workers for text-annotation tasks. *PNAS*, 120(30), e2305016120. doi:10.1073/pnas.2305016120

Hill C, Dahil A, Simpson G, et al. (2026). Large language models for thematic analysis in healthcare research: a blinded mixed-methods comparison with human analysts. *PLOS Digital Health*, 5(4), e0001189. doi:10.1371/journal.pdig.0001189

Leas EC, Ayers JW, Desai N, et al. (2024). Using large language models to support content analysis: a case study of ChatGPT for adverse event detection. *JMIR*, 26, e52499. doi:10.2196/52499

CC BY NC SA

References (2/2)

Spitale G (2025). *Doing research with AI* [lecture slides]. IBME, University of Zurich, 10 December 2025. giovannispitale.net

Spitale G (2026). *AI governance* [bootcamp slides]. DSI Excellence Bootcamp, University of Zurich, 3 September 2026. giovannispitale.net/ai-governance-dsi-excellence-bootcamp

Spitale G, Germani F (2026). Misinformation, disinformation, and autonomy: against the epistemic paternalism trap [preprint]. *Zenodo*. doi:10.5281/zenodo.20321941

Spitale G, Germani F (2026). AI in Science: The 42 problem in AI-assisted research. *eLife*. doi:10.7554/eLife.113137

UZH (2021). Weisung zur Klassifizierung von Informationen (in force since 1 September 2023). rud.uzh.ch

UZH Zentrale Informatik (2026). Übersicht KI-Tools und Services, version of 29 September 2026. zi.uzh.ch

van den Berkmortel B, Westra D, Gifford R, van de Baan FC (2026). Analyzing nurse sentiment in a Dutch regional newspaper during the COVID-19 pandemic: comparative content analysis of GPT-5, Gemini 2.5 Pro, and human coding. *JMIR*, 28, e98374. doi:10.2196/98374

Exercise E: the outputs, the four verification reports and the references they cite are in the exercise pack and key; five of the papers listed (Balt, Gilardi, Hill, Leas, van den Berkmortel) are the sources Output 3 was restricted to.

CC BY NC SA



Let’s keep talking

giovanni.spitale@ibme.uzh.ch

Slides and materials

Information Technology & Experimental Ethics

University of Zurich[UZH]

Institute of Biomedical Ethics and History of Medicine

CC BY NC SA