Briefing

Welcome to Your Vademecum! / 1

What is the vademecum?

We designed this vademecum to be your companion for inquiring into your professional use of Large Language Models (LLMs). Inside, you will find a series of exercises aimed at helping you better understand your LLM practices and how they fit, reshape, or threaten your professional environment. As a physical, analogue resource, it deliberately stands apart from your digital work environment —AIs included—to create a breathing space inside your day-to-day practices. We hope this will encourage you to pause, take a step back, and reflect on your LLM use.

This vademecum is now yours—or, at least, it will become yours over the course of the following weeks. You will fill it with your observations, descriptions, thoughts, and feelings about LLMs. Over time, it will evolve into a personalised journal, chronicling your own journey with these tools. If you're utterly inexperienced with LLMs, feel free to experiment at your own pace. The first two exercises require minimal usage, giving you time to ease into a more regular usage of LLMs. However, before moving on to Exercise 03, you should aim to have had around fifty conversations with the LLM of your choice.

The Ecology of the Vademecum

Because we care about understanding LLMs within their real-world usage and professional ecologies, we think it's only fair to provide an ecological overview of this vademecum's genesis. It originated from the Ecologies of LLMs Practices (EL2MP) research project, where we work with several cohorts of professionals from the same discipline, with each cohort lasting 6 to 8 months.Every cohort completed the same exercises, and met regularly (10 to 15 times) for researcher-led workshops and individual interviews. We later scanned and collected each cohort's vademecum to form our research corpus .

The version that is currently in your hands is a distilled product of those experiments. While it was first created for our research protocols, we've reworked it so that it can be used to find meaning and purpose in all sorts of settings: personal study, academic research, groups of colleagues, or workers' communities—wherever you want to deepen your insight into your LLM practices.

What Can You Do With It?

The vademecum is a modular object. It is bound by two metal rods, allowing you to add, remove, or rearrange pages as needed. The metal rods are secured by small metal pieces; to open it, simply slide these pieces out of the way and unfold the long metal strips. When you're done rearranging the contents, slide the pages back onto the rods, fold the rods, and secure them by sliding the metal pieces back into place.

You can customise the vademecum to fit your needs, adding extra notes, drawings, or other materials as necessary – you'll find extra blank pages at the end. This is a space for documenting everything you experience, from your immediate reactions to the deeper insights that build up over time. Feel free to draw, write, colour, or use whatever means you like to capture your thoughts and observations.

The exercises are gathered into blocks, each one exploring a key question about LLM use. Some exercises are built upon previous ones, so feel free to adapt or skip some—but be aware that skipping may require a bit of creativity to fill in any gaps.

Working With Others

We've found the vademecum is most valuable when it's shared or discussed collectively. You can simply talk about your results with a friend or colleague, or form a larger group where everyone works through the vademecum at the same pace.

Feedback and Data Contribution

If you'd like to share a part of or your entirely completed vademecum with us to support our research or exchange ideas, please get in touch. We're also happy to answer questions, discuss the project more in-depth, or hear your feedback and stories about working with the vademecum.

Becoming a Co-inquirer

Description

Now that you have a sense of how the vademecum works, let's talk about a concept that is at its core: becoming a co-inquirer. The meaning we attribute to the term "co-inquirer" is someone who not only uses LLMs but also actively observes, describes and interprets their own experience. Becoming a co-inquirer is an iterative process. This page serves as a guide for documenting your interactions with LLMs. Refer back to it regularly.

→ *Your role is to document your use of LLMs and help interpret them for an outsider, who knows nothing of your profession.* You regularly take time to step back and reflect on your current and past practices. Why did you do what you did? Why did you do it that way? How did you feel at the time? What were you thinking? Did you act without thinking? To say it concretely, you speak in the first person. Your accounts are not impersonal descriptions of how an AI works; rather, they shed light on what happens between you and the keyboard. Clarify implicit assumptions whenever possible.

→ *You give LLMs a chance.* If you have not already, try incorporating LLMs into your daily work practices by using them regularly throughout the day and week. To criticise or praise LLMs, one should give them a fair trial. Regardless of your current stance on them, you are willing to navigate uncertainty. Pay attention to any discomfort that may arise and let it guide your inquiry. Your role is not to pass judgment on LLMs but to explore how they fit—or don't fit—into what you usually do. What assumptions do you hold? How do the experiments challenge or reinforce your beliefs?

→ *You are both a participant and an observer*. You simultaneously participate in the exercices and reflexively observe your own practices. Let comparison guide your reflection: is there any gap between your practices and those of your peers? Do some exercises go against the flow of your own practices or do they intensify them?

→ *You write accounts grounded in practice*. You don't write general essays. Your starting point is the documented practices. You seek to contextualise, understand, and explain them. Your accounts can be risky, creative, humorous, or cynical—as long as they stem from your honest study of practices. Importantly, you don't try to look your best. These experiments are not academic exercises. There are no grades and no anticipated results. There are no right or wrong answers – only good accounts of what you did.

→ *You pay attention to the mundane and the extraordinary*. Do not assume that you know what the researchers will be interested in. We may be interested in what you find boring, bland, or ordinary. We may be interested in what you find bizarre, unspeakable, irrational, ill-conceived, or poorly expressed. Do not censor your accounts.

How can we reframe the role of LLMs in ordinary work practices?

The Ecologies of LLM Practices (EL2MP) project creates a space of inquiry to document the role of LLMs in various professional practice(s) and the consequences arising from their use.

Context

The booming rise of large language models (LLMs) such as ChatGPT has sparked a rush to produce discourse about these technologies. The quick crystallisation of a shared outlook around a few key themes has narrowed the scope of potential interrogations. Public and scientific debates focus on technical issues: algorithmic bias, confabulation, and intellectual property violations. However, the problems and consequences associated with their actual use – for both their users and their professional contexts – remain largely unexplored. This asymmetry fuels a mechanical view of technological development and its effects, as if the technical analysis of these systems were enough to predict their social impact. Moreover, these discourses present AI as a monolithic and disruptive entity, dismissing the possibility that it may be aligned with existing practices and that its effects may vary depending on situations encountered in one's job.

There is thus an urgent need to move beyond predictions about the future of work, to account for the professional contexts in which LLMs are used, and to identify current issues – not prospective ones. How do AI's well-known problems (bias, confabulation, etc.) manifest in established practices? What new, unexpected problems are surfacing? How do LLMs shape individual work practices? And, in turn, how do professional environments shape LLMs and their use?

Objective

To answer these questions, the Ecologies of LLM Practices (EL2MP) project creates research areas dedicated to workers for them to document and reflect on their use of LLMs. Our investigation aims to highlight the "savoir-faire", expertise, and values of workers rather than those of AI designers or economic decision-makers. The project aims to scrupulously examine how LLMs fit into various professional practices. EL2MP will investigate how users relate to LLMs in terms of:

→ *Evaluation:* How do professionals assess the value LLMs add or remove?
→ *Effort:* What new kinds of work do LLMs require from their users?
→ *Perception:* How do workers evaluate and perceive LLMs over time as they continue working with them?

Methodology

To carry out the project, we developed an experimental research protocol based on the active involvement of participants. The fruit of a collaboration between sociologists, science and technology researchers (STS), and designers, this protocol aims to create:

→ *Room for hesitation:* Through a series of exercises, the protocol provides participants with various means of documenting and reflecting on their use of LLMs. We aim to establish a framework where doubt and hesitation are not only accepted but encouraged. Some exercises introduce deliberate pauses in professional routines, causing participants to take a step back from their own practices: during these pauses, judgment is suspended, giving users enough room to question their instinctive habits and feel unsure about what they once thought was certain. Other exercises aim to intensify the use of LLMs to shed light on what would otherwise remain too subtle to be perceived.

→ *An ecological archive of practices:* The digital traces from LLM use serve as the starting point for the exercises we designed. They are analysed, contextualised, and discussed individually and collectively. Through this process, participants progressively build a multimodal archive (audio, photos, videos, drawings, and logbooks) that tangibly reflects their LLMs experience and how they are incorporated into various professional norms and material configurations.

→ *A workbook (vademecum):* The workbook collects all the exercises and their instructions, organized into thematic blocks. Designed as a modular object, where pages can be added or removed as needed, it accompanies each participant throughout the protocol. The vademecum serves as both the material support for the protocol and the project's means of dissemination, as it can be reproduced and adapted in various academic and professional contexts.

Research team

→ Donato Ricci: Principal Investigator, Designer
→ Gabriel Alcaras: Postdoctoral researcher, Sociologist
→ Tommaso Prinetti: Research Assistant, Designer
→ Zoé de Vries: Research Assistant, Linguist

Qualifying

Where do LLMs live in our professional practices, norms, and relationships?

The Qualifying Block serves as your entry point in the core of the vademecum – the exercises. The tasks in this first block share a common goal: to help you evaluate where you currently stand in your professional use of LLMs.

The exercises in the Qualifying Block encourage you to explore your mental representations of LLMs and account for your actual LLM and work practices. Some also aim to broaden your perspective by acting as a stepping stone towards diversifying your LLM usage, enabling you to rethink and expand the ways in which you integrate these tools into your professional activities.

EX*1* – Draw It Like You See It

Instructions

In this exercise, you are asked to sketch two different mental representations: one depicting how you visualize the functioning of an LLM and the other depicting how you visualize your work environment. No need to do additional research for these sketches, just draw off the top of your head based on your own perceptions.

You have no limits other than the square boundary for your drawings.
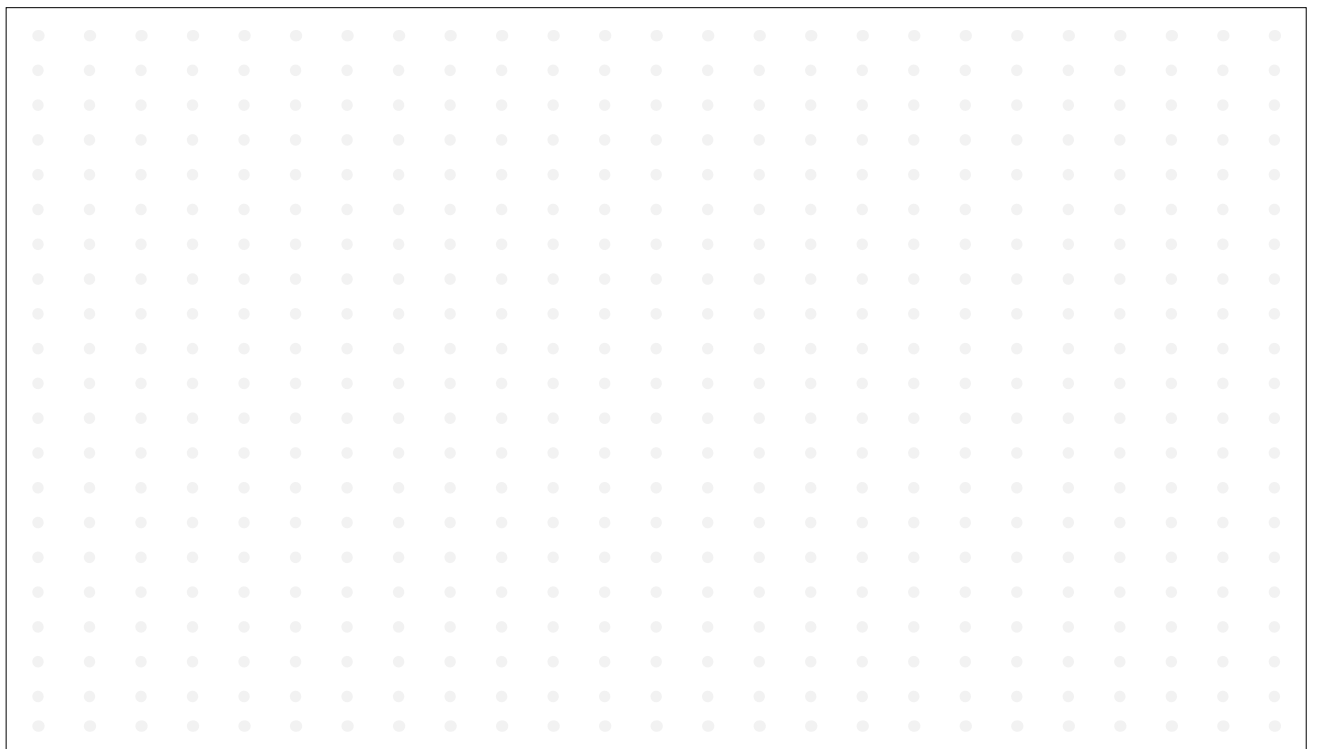
EX*1* – Draw It Like You See It

In a few words, what do you do?

_____

_____

What does your professional environment look like?

What does an LLM look like?

EX^2 - *Harvesting Tasks*

In this exercise, you will conceive two lists. The first captures the range of tasks you currently perform in your professional practice, and the second imagines additional tasks you could perform with the help of an LLM. It is best to perform this exercise in the span of multiple sessions, giving yourself time to think and add new ideas.

→

EX²ᵃ – All the Things You Do

Instructions

Write a list of tasks related to your professional practice. You probably perform many, many more tasks in your job than you actually realize. Doing this exercise in several sittings rather than just one is preferable to have a clearer picture of your task range.

You can use the questions below to trigger your memory:

→ What tasks do you dislike doing? Which ones do you enjoy?

→ What tasks do you have to do regularly? Sometimes? Rarely?

→ What are some tasks that you regularly procrastinate on? Are there optional tasks that you still tend to do very often?

→ In your field or profession, what tasks are basic? Necessary but tedious? Which ones are the mark of a good professional or an expert?

→ Which tasks take the most time? Which ones cost you the most money?

The "Task ID" allows you to identify your task without having to rewrite its longer description every time. This will be useful in later exercises when you will have to refer back to these tasks.

EX*2a* – All the Things You Do

Task list

Task ID: X5   Task:

Task ID: S8   Task:

Task ID: E9   Task:

Task ID: H4   Task:

Task ID: B5   Task:

Task ID: U1   Task:

Task ID: F0   Task:

Task ID: I0   Task:

Task ID: S6   Task:

Task ID: S0   Task:

Task ID: H9   Task:

Task ID: B3   Task:

Task ID: Q1   Task:

Task ID: D2   Task:

Task ID: A3   Task:

Task ID: Z7   Task:

Task ID: C9   Task:

EX^{2a} – All the Things You Do

Task list

Task ID: A7   Task:

Task ID: K7   Task:

Task ID: G1   Task:

Task ID: X2   Task:

Task ID: Z5   Task:

Task ID: R0   Task:

Task ID: L6   Task:

Task ID: N5   Task:

Task ID: P0   Task:

Task ID: M8   Task:

Task ID: Q0   Task:

Task ID: B0   Task:

Task ID: Q9   Task:

Task ID: V5   Task:

Task ID: G8   Task:

Task ID: I1   Task:

Task ID: R1   Task:

EX*2b* – All the Things You Could Do

Instructions

In this exercise, think about the tasks you currently do without the help of LLMs but where you could benefit from their assistance.

In other words, this list is not about automation (delegating to the machine) but augmentation (doing more tasks or diversifying tasks). Think about activities or projects you have not attempted yet or tasks you might hesitate to delegate.

Even if you suspect that an LLM might perform poorly on a given task, include it in your list; this exercise is about exploring potential opportunities rather than evaluating performance.

EX^{2b} – All the Things You Could Do

Task list

Task ID: K0   Task:

Task ID: M4   Task:

Task ID: M7   Task:

Task ID: M0   Task:

Task ID: J6   Task:

Task ID: Z6   Task:

Task ID: O2   Task:

Task ID: C3   Task:

Task ID: L4   Task:

Task ID: K4   Task:

Task ID: H3   Task:

Task ID: G4   Task:

Task ID: T4   Task:

Task ID: T1   Task:

Task ID: B1   Task:

Task ID: H6   Task:

Task ID: G2   Task:

EX^{2b} – All the Things You Could Do

Task list

Task ID: T2   Task:

Task ID: N3   Task:

Task ID: J3   Task:

Task ID: I8   Task:

Task ID: Z4   Task:

Task ID: E4   Task:

Task ID: I6   Task:

Task ID: C5   Task:

Task ID: B2   Task:

Task ID: F9   Task:

Task ID: U9   Task:

Task ID: I9   Task:

Task ID: K9   Task:

Task ID: A1   Task:

Task ID: F5   Task:

Task ID: P3   Task:

Task ID: O6   Task:

EX^3 – Taking stock

In this exercise, you will review the history of your interactions with your LLM (e.g., ChatGPT) to take stock of your practices. You will capture your overall impressions, identify patterns in your usage, and note memorable interactions.

This exercise will be most useful once you have gained some experience using LLMs. If you still lack exposure to them, meaning that you have had fewer than 50 conversations, we recommend that you revisit it later.

→

EX³ᵃ – Rough Impressions

Instructions

Set a timer and spend 5 minutes reflecting on your overall LLM use. Do not look at your conversation history yet!

Assess which tasks you delegate and assign a rough percentage to each (e.g., one-third of my tasks are translation, and the rest are drafting email answers). Consider how frequently you use the LLM (daily, weekly, etc.).

Reflect on the structure of your interactions and decide how many of your conversations are: short exchanges (a single request and answer), longer conversations (several iterations), or extended dialogue (conversations that last more than a day).

Finally, recall some memorable moments in your use of the LLM. Think back to two or three sessions that left a lasting impression.

Describe them briefly and why they were memorable.

EX³ᵃ – Rough Impressions

Draw & sketch here

Write notes here

EX^{3b} – Rolling the Dice

Instructions

Having taken stock of these rough impressions, you will now index your past interactions with the LLM of your choice. We assume that you have had at least fifty conversations with an LLM, which you will review from the most recent to the oldest using a dice. This will spare you from having to read through your entire history in detail.

Start by opening your primary LLM interface and locating your conversation history. The conversation titles should serve as a memory aid, but if the title is insufficient to jog your memory, click on the conversation and open it.

1.   First, roll the dice.

Number (X) indicates how many conversation titles you should review and how many conversations you should skip after reviewing the first X conversations.

→   Let's say you roll a six: you will review the first six conversations, then skip the next six, and roll again.

2.   Review the conversations according to the instructions below.

3.   Continue this process until you have covered your entire conversation history.

Assign each conversation a task category; if the conversation represents a new type of task, create a new category and tick off a box in the gauge. If it already corresponds to an existing category, simply tick off another box in that category.

→   Additionally, if a task stands out, record its title in your "Memorable Conversations" list.

You can record a maximum of eight memorable conversations. A conversation can be memorable for various reasons: it might provoke a specific emotion (such as fear, excitement, frustration, or amusement); the LLM may have performed exceptionally well or poorly; it may have triggered a moment of moral or ethical dilemma; it could have shifted your perception of LLMs; or it might have had real consequences on your work, among other reasons.

Example

Six conversations belong in the "translation" category.

Category:	Translation
Occurrences:	

EX³ᵇ – Rolling the Dice / 1

Task categories

Category:

Occurrences:

Category:

Occurrences:

Category:

Occurrences:

Category:

Occurrences:

Category:

Occurrences:

Category:

Occurrences:

Catégorie:

Occurrences:

Category:

Occurrences:

Category:

Occurrences:

Category:

Occurrences:

Category:

Occurrences:

EX^{3b} – Rolling the Dice / 2

Task categories

Category:

Occurrences:

Category:

Occurrences:

Category:

Occurrences:

Category:

Occurrences:

Category:

Occurrences:

Category:

Occurrences:

Memorable conversations

Conversation title:

Conversation title:

Conversation title:

Conversation title:

Conversation title:

Conversation title:

Conversation title:

Conversation title:

EX³ᶜ – The *Story* Told By Your LLM

Instructions

Looking back at your history with LLMs, answer the following open-ended questions based on your conversation history..

EX³ᶜ – The *Story* Told By Your LLM / 1

What do your LLM conversations capture accurately about your work?

Reflect on the narrative they shape and the stories they tell about your professional practices, routines, and strategies.

EX^{3c} – The *Story* Told By Your LLM / 2

If someone were to look solely at your LLM conversations to understand your work, what aspects would be overlooked or misrepresented?

Reflect on the shadows, missing context, and the work that happens before and after using the LLM that these conversations might not capture.

EX*3c* – The *Story* Told By Your LLM / 3

What has the machine done for you?

Describe how the LLM has been helpful and valuable to your work..

EX³ᶜ – The *Story* Told By Your LLM / 4

What have you done for the machine?

Reflect on the work you've provided and the time you've invested to get
good results out of the LLM.

EX⁴ – *Memorable* Conversations

Instructions

This exercise invites you to revisit conversations that have left a lasting impression. To refresh your memory, refer to the conversation titles you noted in Ex03b ("Rolling the Dice") and open them in your LLM interface. If you lack examples, you can return to your conversation history and add new conversations if necessary.

Start by copying the conversation title and briefly describing its content. Then, explain why the conversation stood out. Reflect on their emotional impact—whether it evoked excitement, frustration, amusement, fear, or any other strong emotion. Consider any moments that raised moral or ethical questions, shifted your views on LLMs, or had significant consequences for your work. Your descriptions should capture both your immediate feelings and the broader implications of the interaction.

Finally, shift your focus to the performance of LLMs. For each conversation, rate the performance of the LLM from one to five stars (one being abysmal and five excellent), and write a brief evaluation of the LLM's performance. Then, explore why you believe the LLM performed as it did.

Exemple

Conversation title:

Translating email

Content:

This conversation was about translating emails for an economic summit in Paris

Why it stood out:

I remember this conversation because GPT perfectly understood the language and word-choice needs at stake for each email it was asked to generate.

Describe the performance:

The LLM did poorly as its answers were very vague. It was also unable to remember things and hallucinated a lot.

Score: 1 (Abyssal) —✗———————————— 5 (Excellent)

EX⁴ - *Memorable* Conversations / *Playground 1*

Conversations

Impact

Performance

Conversation 1

Conversation title:

Content:

Why it stood out:

Describe the performance

Score : 1 (Abyssal) ——————— 5 (Excellent)

Conversation 2

Conversation title:

Content:

Why it stood out:

Describe the performance

Score : 1 (Abyssal) ——————— 5 (Excellent)

Conversation 3

Conversation title:

Content:

Why it stood out:

Describe the performance

Score : 1 (Abyssal) ——————— 5 (Excellent)

Conversation 4

Conversation title:

Content:

Why it stood out:

Describe the performance

Score : 1 (Abyssal) ——————— 5 (Excellent)

EX⁴ - *Memorable*
Conversations / *Playground 1*

EX⁴ - *Memorable* Conversations / *Playground 2*

Conversations

Impact

Performance

Row 1

Conversation title:

Content:

Why it stood out:

Describe the performance

Score : 1 (Abyssal) ———————— 5 (Excellent)

Row 2

Conversation title:

Content:

Why it stood out:

Describe the performance

Score : 1 (Abyssal) ———————— 5 (Excellent)

Row 3

Conversation title:

Content:

Why it stood out:

Describe the performance

Score : 1 (Abyssal) ———————— 5 (Excellent)

Row 4

Conversation title:

Content:

Why it stood out:

Describe the performance

Score : 1 (Abyssal) ———————— 5 (Excellent)

EX^4 - *Memorable Conversations* / *Playground 2*

EX^5 - *Subtracting* the Machine

To talk about LLMs, we rely on words like «machine,» «AI,» or «ChatGPT.» While convenient, these terms exempt us from thinking about the role of this «other.» In this exercise, we aim to clarify this relationship by removing these common labels from our language and coming up with new terms.

→

EX^{5a} – A *Thought* Experiment

Instructions

In our everyday interactions with LLMs, we naturally tend to attribute our questions and the resulting answers to a dialogue between "ourself" and the machine.

→ We might say, "I asked ChatGPT to write an email," or "I asked it to generate a summary." In this exercise, we invite you to peel back one layer of that dialogue by subtracting the notion that the answer comes from a machine.

What remains when you remove the labels "machine," "AI," "LLM," "ChatGPT," or even "it" from your descriptions?

Revisit your most memorable conversations with the LLM and examine them closely. For each conversation, ask yourself: Who is asking, and who is answering? Choose ten different interactions— each consisting of your message paired with the machine's response. To anchor your memory, extract and note an excerpt from each interaction, such as the first five words of your message and the machine's response.

Instead of describing the interaction with usual terms, choose alternative words or roles. You might decide that your prompt could come from "a manager" or "a demanding client," while the response from the machine could be characterized as coming from "an advisor," "a colleague," or "a consultant." or even "a ghost."

→ You are free to choose any words except the subtracted words: this could be a professional title ("an assistant"), an online service ("Wikipedia"), a corporation ("McKinsey"), or even an undefined entity ("a ghost").

Work through these interactions iteratively. After your initial substitutions, evaluate whether the new labels accurately capture the essence of each exchange. If your descriptions feel too generic or are repeated across several interactions, refine them with another round of subtraction. The aim is not to arrive at the "perfect" alternative but to explore how the re-description of your interactions transforms your perception of both your role and that of the system.

Example

Conversation title:

USER: A company happiness manager

USER: An event planner

USER: A private individual

User:
Can you give me easy-to-make dish ideas for a birthday celebration with 60 people?

LLM:
For a birthday celebration with 60 people, it's best to go with dishes that are simple to make in large quantities. Here are some easy-to-make dish ideas

LLM: A friend with a knack for cooking

LLM: A cooking website

LLM: A caterer

EX*5a* – A *Thought* Experiment / 1

Conversation title:

USER[3]:

USER[2]:

USER[1]:

User:

LLM:

LLM[1]:

LLM[2]:

LLM[3]:

Conversation title:

USER[3]:

USER[2]:

USER[1]:

User:

LLM:

LLM[1]:

LLM[2]:

LLM[3]:

EX^{5a} – A *Thought* Experiment / 2

Conversation title:

USER³:

USER²:

USER¹:

User:

LLM:

LLM¹:

LLM²:

LLM³:

Conversation title:

USER³:

USER²:

USER¹:

User:

LLM:

LLM¹:

LLM²:

LLM³:

EX^{5b} – What *Remains* After Subtraction

Instructions

You have now built a lexicon through a series of term subtractions. What do you think of the final terms? Do they adequately describe who is writing/speaking? Do some feel uncanny or weird to use? Why? Consider whether the way you've redefined the speakers matches the tone of the interaction. For example, does the response read as if it were actually coming from a professional caterer?

Important

> Write your thoughts on the following page

EX^{5b} – What *Remains* After Subtraction

Write here:

Benchmarking

How do we assess the value of an LLM?

The Benchmark Block guides you through the process of designing a tailored AI trial using LLMs to test their relevance and usefulness in your professional activities. The exercises challenge you to identify and select key tasks that represent the core of your work. By sorting tasks, reflecting on their professional significance, and evaluating whether they could benefit from the integration of LLMs, you will shape a comprehensive trial.

In addition to creating a personal benchmark, this block offers you the opportunity to play and experiment with different LLMs, evaluate the extent to which they vary from one another, and determine if one of them is better suited to your professional practices.

EX⁶ – *Design* Your AI Trial

The four core tasks you select in this exercise are crucial because they represent the very heart of your professional practice. They are not merely arbitrary tasks but carefully chosen activities that accurately reflect your work. By isolating these core four tasks, you create a focused, meaningful AI trial that can reveal both the strengths and the limitations of an LLM in handling the demands of your daily work.

The exercise is structured in three different parts: first, sorting all your potential tasks into a Venn diagram to categorize what you have to do versus what you enjoy doing; second, choosing the four tasks that best capture the diversity and significance of your work; and finally, reflecting on these choices to understand what each task reveals about your professional practices and what is at stake in delegating them to an AI.
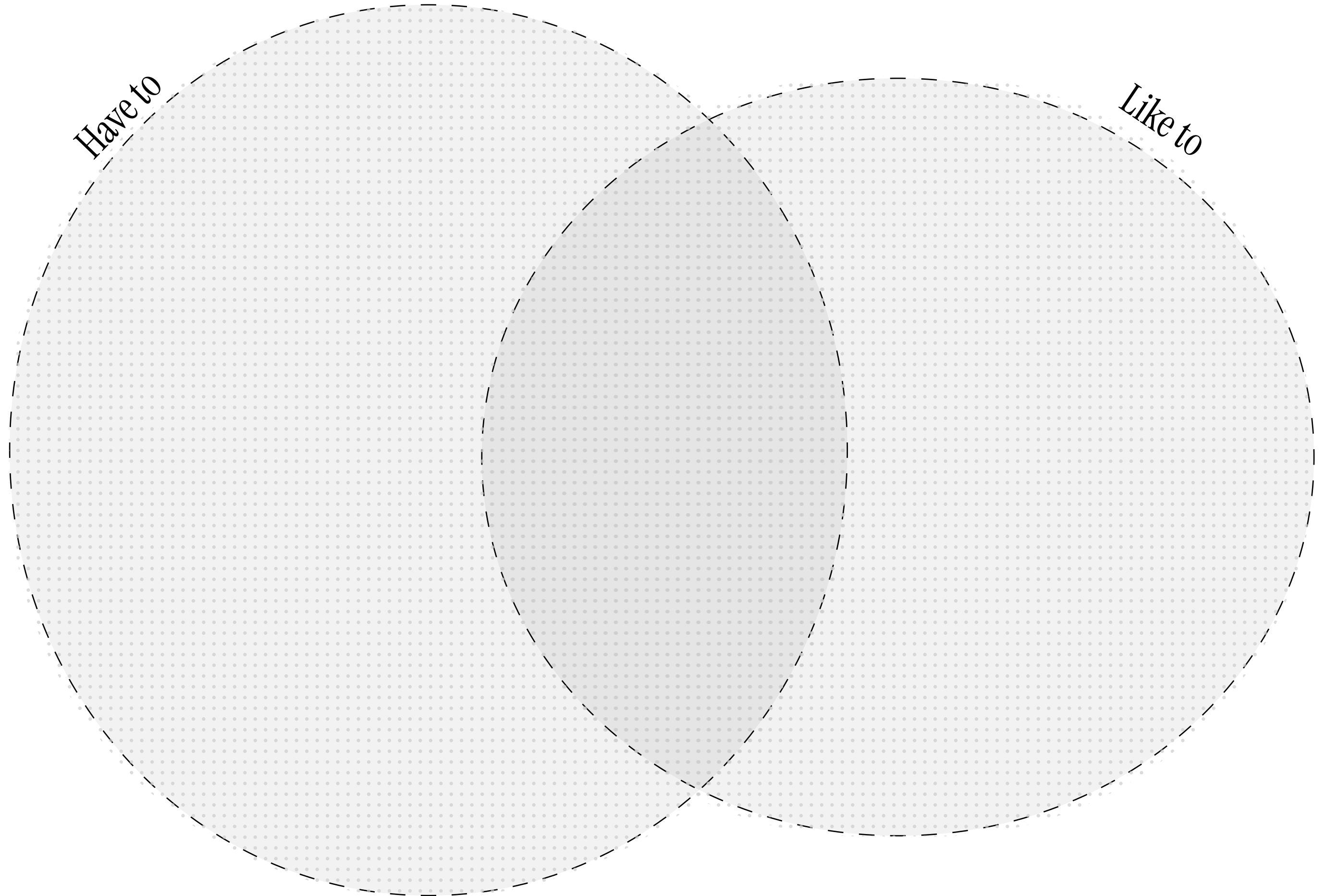
→

Have to

Like to

EX^{6a} - *Business* and/or Pleasure?

Instructions

In this exercise you will sort the tasks you currently do (and those you might do in the future, refer to the "Harvesting Tasks" exercise) into three overlapping categories:

→ The tasks you must perform for your job (obligations),
→ The tasks you enjoy doing,
→ The tasks that fall in between.

Complete the Venn diagram below using the task identifiers from Exercise X (for example, AA1, BB2).

EX*6b* – Choosing Your *Core* Four

Instructions

Now, select four tasks to serve as the basis of your AI trial. Consider these three criteria:

→ *Choose tasks that matter professionally.* These tasks should reflect your professional practices, whether they are frequent or rare.

→ *Pick a diverse range of tasks.* Choose at least one task from each area of the Ex06a diagram (have to, like to, both). Include at least one (ideally two) task you would usually hesitate to delegate to an LLM.

→ *Select tasks fairly for the LLM.* You are allowed to choose a task on which you suspect the LLM will perform poorly. However, avoid including too many tasks that would be excessively challenging even for a skilled professional.

Keep in mind that these four tasks form the basis for selecting the model that best aligns with your work practices. Once you have made your selection, explain your reasoning. Describe how you feel about your choices and why you made them.

EX*6b* – Choosing Your *Core* Four

Identifier le carré d'as

1st Task ID : [ ] | Task: (Have to) (Like to) (Delegation: (///) (Hesitant))

2nd Task ID : [ ] | Task: (Have to) (Like to) (Delegation: (///) (Hesitant))

3rd Task ID : [ ] | Task: (Have to) (Like to) (Delegation: (///) (Hesitant))

4th Task ID : [ ] | Task: (Have to) (Like to) (Delegation: (///) (Hesitant))

Judging the core four

EX^{6c} – What's On The *Line* ?

Task ID:

Task fit:

In what way does this task reflect a core part of your job or expertise? Is it something you do regularly or only on occasion? Is it a boring or an exciting task? ?

Professional relevance:

Does the task highlight any specific skills or crucial professional expertise? Are there any industry norms, professional codes, or expectations that this task should respect?

Personal outcome:

How will evaluating the LLM on this task guide you toward the model that suits your work best?

Rating the result:

Does this task need to be perfect, or is there room for "good enough"? How will you decide if the LLM's performance is sufficient to support you professionally?

Failure

What would be a deal-breaker in your field?

Write here:

Good enough

What would be a usable or helpful result?

Write here:

Success

What would be a perfect result?

Write here:

Assessing your expectations:

How well do you think the LLM will perform on this task?
Circle the relevant option.

| Terrible | Pretty bad | Good enough | Excellent |

How sure are you of that answer?
Circle the relevant option.

| Completely unsure | Unsure | Pretty sure | Confident |

EX*6c* – What's On The *Line* ?
/ *Playground 1*

Instructions

Now that you've designed your trial, let's step back and reflect on what each task reveals about your professional practices. This exercise will help you explore how these tasks connect to your work and what's on the line in your trial.

EX^{6c} – What's On The *Line* ?

Task ID:

Task fit:

In what way does this task reflect a core part of your job or expertise? Is it something you do regularly or only on occasion? Is it a boring or an exciting task? ?

Professional relevance:

Does the task highlight any specific skills or crucial professional expertise? Are there any industry norms, professional codes, or expectations that this task should respect?

Personal outcome:

How will evaluating the LLM on this task guide you toward the model that suits your work best?

Rating the result:

Does this task need to be perfect, or is there room for "good enough"? How will you decide if the LLM's performance is sufficient to support you professionally?

Failure

What would be a deal-breaker in your field?

Write here:

Good enough

What would be a usable or helpful result?

Write here:

Success

What would be a perfect result?

Write here:

Assessing your expectations:

How well do you think the LLM will perform on this task?
Circle the relevant option.

Terrible	Pretty bad	Good enough	Excellent

How sure are you of that answer?
Circle the relevant option.

Completely unsure	Unsure	Pretty sure	Confident

EX^{6c} - What's On The *Line* ?
/ *Playground 2*

EX*6c* – What's On The *Line* ?

Task ID:

Task fit:

In what way does this task reflect a core part of your job or expertise? Is it something you do regularly or only on occasion? Is it a boring or an exciting task? ?

Professional relevance:

Does the task highlight any specific skills or crucial professional expertise? Are there any industry norms, professional codes, or expectations that this task should respect?

Personal outcome:

How will evaluating the LLM on this task guide you toward the model that suits your work best?

Rating the result:

Does this task need to be perfect, or is there room for "good enough"? How will you decide if the LLM's performance is sufficient to support you professionally?

Failure

What would be a deal-breaker in your field?

Write here:

Good enough

What would be a usable or helpful result?

Write here:

Success

What would be a perfect result?

Write here:

Assessing your expectations:

How well do you think the LLM will perform on this task?
Circle the relevant option.

| Terrible | Pretty bad | Good enough | Excellent |

How sure are you of that answer?
Circle the relevant option.

| Completely unsure | Unsure | Pretty sure | Confident |

EX⁶ᶜ – What's On The *Line* ?
/ *Playground 3*

EX⁶ᶜ – What's On The *Line* ?

Task ID:

Task fit:

In what way does this task reflect a core part of your job or expertise? Is it something you do regularly or only on occasion? Is it a boring or an exciting task? ?

Professional relevance:

Does the task highlight any specific skills or crucial professional exper-tise? Are there any industry norms, professional codes, or expectations that this task should respect?

Personal outcome:

How will evaluating the LLM on this task guide you toward the model that suits your work best?

Rating the result:

Does this task need to be perfect, or is there room for "good enough"? How will you decide if the LLM's per-formance is sufficient to support you professionally?

Failure

What would be a deal-breaker in your field?

Write here:

Good enough

What would be a usable or helpful result?

Write here:

Success

What would be a perfect result?

Write here:

Assessing your expectations:

How well do you think the LLM will perform on this task?
Circle the relevant option.

| Terrible | Pretty bad | Good enough | Excellent |

How sure are you of that answer?
Circle the relevant option.

| Completely unsure | Unsure | Pretty sure | Confident |

EX*6c* – What's On The *Line* ?
/ *Playground 4*

EX*7* – Preparing for The Trial

In this exercise, you will detail the tasks and select the AI models, laying the ground-work for your trial.

→

EX⁷ᵃ – Writing *Guidelines*

Instructions

Refer back to the four core tasks you have selected. You will now write detailed instructions for each of them. Remember, they will be input verbatim into the LLMs, and you will not have the opportunity to provide feedback to iterate.

Choose your words carefully, and include as much context or information as necessary.

You may also include documents (PDFs, images, text documents, and so on) if your task requires it..

→   If it helps, consider these instructions a detailed email with attachments you would write to a colleague who knows nothing of your job but who possesses sufficient skills and eagerness to realize the task adequately.

Notice

Do this exercise in the digital format, as you will need to copy / paste your instructions to send them out in Exercise 8.

EX⁷ᵇ – *Choosing* the Models

Instructions

You have now chosen four core tasks and written detailed instructions for each. Four diffe-rent models will execute your core four tasks. The main three models are ChatGPT, Claude, and Mistral.

You may choose what you want to test for the fourth model.

You can choose either:

→ Another general-purpose model, like DeepSeek, Llama 3, ...

→ Any domain-specific model that better suits your professional needs (journalism, law, etc).

→ A fellow co-inquirer, a good friend, or a trusted colleague whom you would ask to per-form the tasks instead of a machine.

Write your choice down here:

	MISTRAL AI_
ChatGPT	✳ Claude

EX^8– Gathering *Evidence*

Instructions

In this exercise, you will enlist the help of a fellow co-inquirer, a trusted friend, or a colleague. This partner will be the one gathering the evidence on your behalf.

Begin by sharing with your partner the list of the four models you selected and ask them to assign each model an animal code — a giraffe , a sloth , a dragon , or a turkey .

These animal names will serve as code names to ensure that when the trial results are returned, you remain unaware of the models' identities and thus maintain impartiality. Next, provide your partner with your list of tasks, including each task's ID, name, detailed instructions you crafted earlier, and any added documents the models might require.

Your partner's role is simply to copy these instructions into the models and then collect the first answer each model provides. If you chose to test a human in the previous exercise, your partner will transfer the detailed instructions to them.

Each answer should be stored under the corresponding task ID and animal code name; for example, if ChatGPT (who was assigned the "sloth") produces a response for a vital email task labeled AA12, the response should be recorded as "AA12/sloth." Since four models execute each task, your partner will eventually return sixteen labeled results.

You will assess the quality of these results in the next exercise.

EX^9 – Judgment Day

In this exercise, you will evaluate the outcomes of your four tasks by judging and ranking the blinded, animal-coded responses provided by the models. Later, you will review these observations to render a final verdict explaining which model performed best. Finally, you will engage in a mock appeal process to critically reassess your trial's design in terms of usefulness and fairness.

→

EX⁹ᵃ – Weighing the *Evidence*

Instructions

Each task has its designated playground and deck of four cards which represent four different results from four models. To ensure a fair judgment, the results will be blinded, and each model will be represented by an animal (for example, a giraffe , a dragon , a sloth , or a turkey).

Start by preparing the materials for the trial. Enter the task IDs on the deck of cards and the playgrounds.

Then, acting as a judge, you will proceed with ranking the four results obtained for each task. Place the four cards on the playground, arranging them from left to right, left representing the worst performance and right being the best.

→ You can use the space between cards to reflect the difference between each performance: cards with equivalent results can be stacked, while a card with a significantly better (or worse) result can be placed far apart from the others.

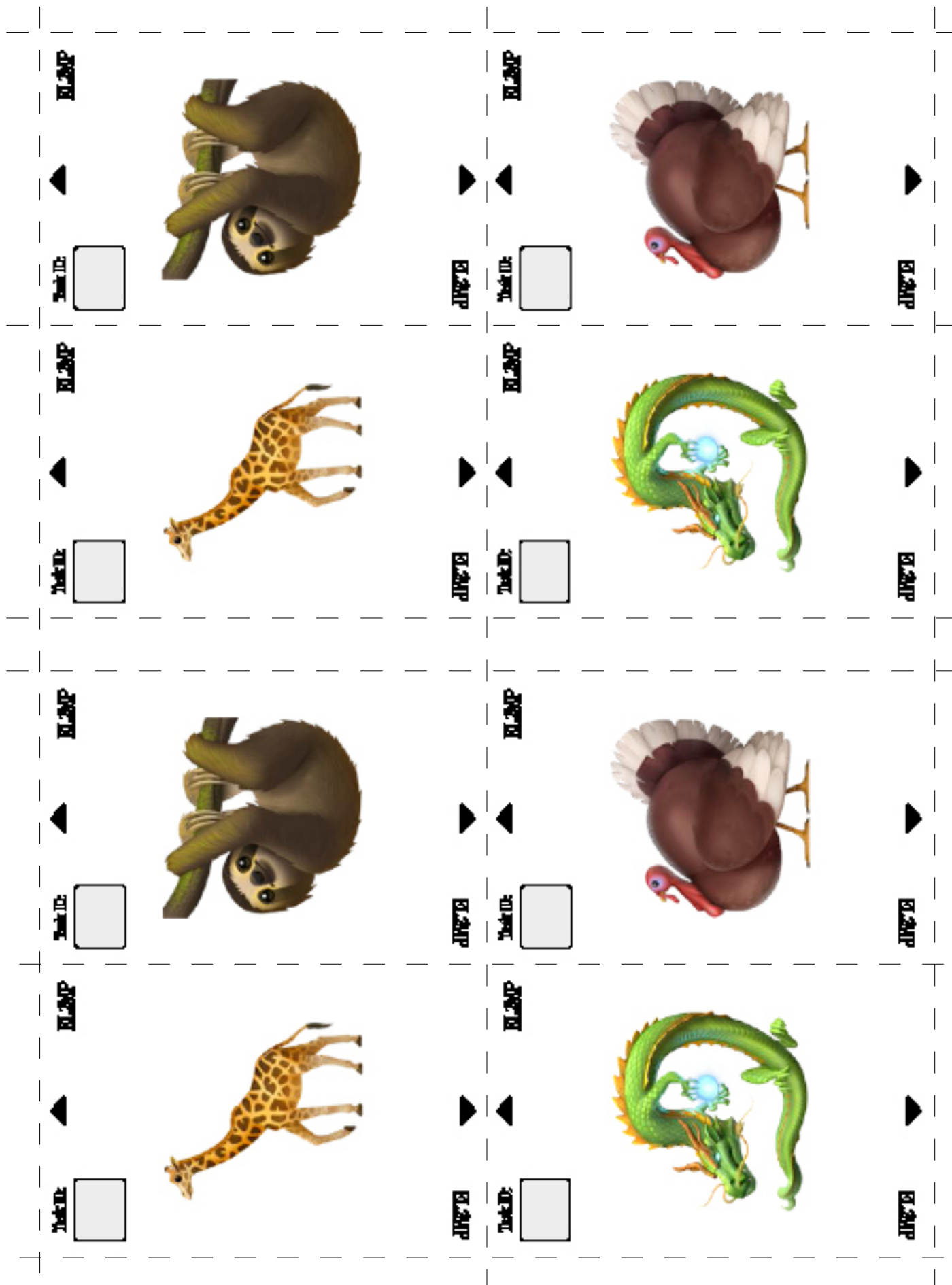Feel free to draw, write or take notes directly on the playgrounds.

You can refer to the scans of playgrounds completed by previous participants for additional guidance.

Exemples
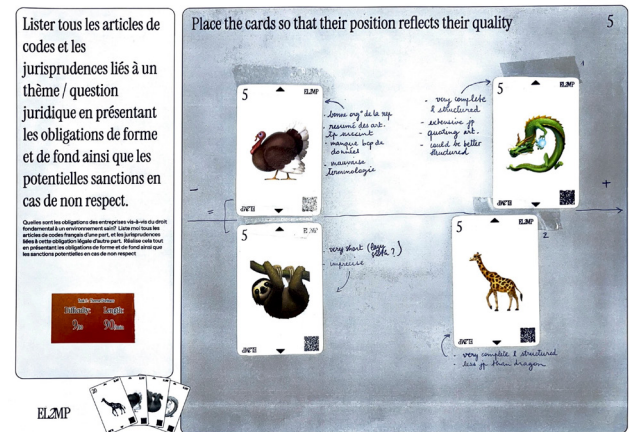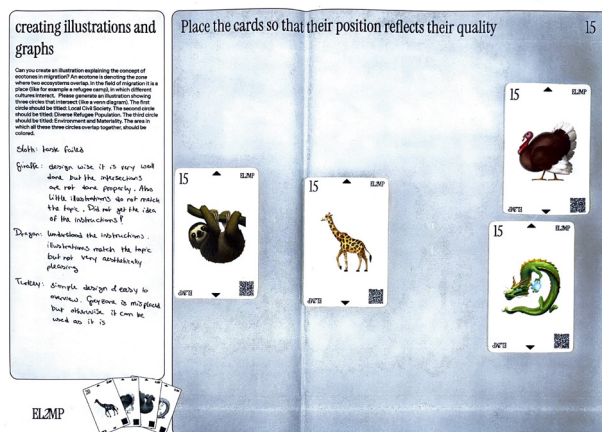
EX⁹ – Cut-A-Card Game

EL2MP

EL2MP

EL2MP

EL2MP

EL2MP

EL2MP

EL2MP

EX^{9a} – Playground

Task ID:

Your notes here:

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Place the cards so that their position reflects their performance

−   _____   +

EX⁹ᵃ – Weighing the *Evidence*
/ *Playground 1*

EX^{9a} – Playground

Task ID:

Your notes here:

Place the cards so that their position reflects their performance

—   +

EX^{9a} – Weighing the *Evidence* / *Playground 2*

EX⁹ᵃ – Playground

Task ID:

Your notes here:

Place the cards so that their position reflects their performance

−   +

EX⁹ᵃ – Weighing the *Evidence*
/ *Playground 3*

EX^{9a} – Playground

Task ID:

Your notes here:

+

Place the cards so that their position reflects their performance

−   +

EX *9a* – Weighing the *Evidence*
/ *Playground 4*

EX⁹ᵇ – The Verdict is In

Instructions

You will now review your observations and render your final judgment. After carefully ranking the responses for each task using the blinded, animal-coded results, write a comprehensive verdict explaining which model, identified by its animal code name, performed best overall and why. Your verdict should be substantiated, including examples from the rankings, and clearly outline the criteria you used to make your judgment.

Important

Écrivez sur la page suivante

EX^{9b} – The Verdict is In

Write here:

EX⁹ᶜ – Revisiting the *Ruling*

Instructions

In this exercise, you will critically evaluate the trial you designed through a mock appeal process. Using the guide below, ask questions about your AI trial. The assessment will focus on two main dimensions:

→ *Usefulness:* assessing whether the trial accurately reflected the professional relevancy of LLM usage.

→ *Fairness:* examining whether the trial was designed in a way that gave the LLM a reasonable chance to perform well.

You will then decide if there are sufficient grounds to appeal your initial ruling on the trial's effectiveness.

EX⁹ᶜ – Revisiting the *Ruling* / 1

Professional Usefulness

Criteria	Useful	Somewhat Useful	Useless
Task Relevance Did the trial include tasks that are representative of your core professional responsibilities?			
Task groundedness Did the trial accurately simulate real-world conditions where you would typically use an LLM?			
Task Impact Were the selected tasks ones that, if successfully delegated to an LLM, would make a substantial difference in your workflow (e.g., by saving significant time, reducing effort, or improving results)?			

Other notes

EX⁹ᶜ – Revisiting the *Ruling* / 2

Fairness Assessment

Criteria	Fair	Somewhat Fair	Unfair
Task Difficulty Were the tasks designed in such a way that the AI had a reasonable chance of performing well, based on its known strengths and limitations?			
Clarity of Instructions Were the instructions for the tasks clear, specific, and consistent enough to allow the AI to perform at its best?			
Task Variety Did the trial include a variety of tasks (in terms of type and complexity) that allowed the AI to demonstrate its abilities across different domains, without favoring tasks inherently better			

Other notes

EX^{9c} – Revisiting the *Ruling* / 3

Final ruling

Put an X on the appropriate box:

Sufficient ground for appeal	Insufficient ground for appeal

What would you change if you had to redo the AI trial?

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Prompting

Does prompting reconfigure our work relationship with the LLM?

The Prompting Block questions what it means to "get better" at using LLMs. You will learn about various prompting techniques, which are standardized strategies designed to yield better or more optimized outcomes. Rather than accepting these techniques at face value, the exercises will encourage you to engage them critically thanks to the learning materials. You will be mindful of their effects on the machine and on yourself.

This block also marks the halfway point in the exercises, so you will get an opportunity to look back on what you've already accomplished and what this journey has meant so far.

EX¹⁰ – The Art of the *Prompt*

Instructions

When learning something new—whether painting a wall, preparing a meal, or crafting an object—we often turn to tutorials: a video, a written guide, or step-by-step instructions. Tutorials promise clarity and simplicity: follow these steps carefully, and the result should be perfect.

Yet, in practice, tutorials rarely deliver exactly what they promised. The wall you painted ends up looking patchy, your dish tastes differently, and your craft project has uneven seams. Why? Perhaps you missed subtle but critical instructions, your tools or materials differed slightly, or the tutorial itself was misleading or incomplete.

→ Working with Large Language Models (LLMs) presents a similar challenge. We are often told that the improvement of our interactions with these tools depends on our crafting better prompts—clearer instructions, more context, or specific phrasing.

But just like with tutorials, improving prompts doesn't always yield the expected outcome. Did your prompt genuinely improve the LLM's response, or did it merely shift your expectations? Was the LLM unable to grasp your intent, or did your request lack precision?

In this exercise, you will experiment with a range of prompting techniques.

You'll test them in your everyday practice, reflect on their effectiveness, and examine how these methods impact not only the LLM's responses but also your own workflow, assumptions and interactions.

EX^{10a} – Room for *Improvement*

Instructions

→ Begin by revisiting your work from the Benchmarking Block.

→ Choose one task where an LLM underperformed or delivered disappointing results.

→ Carefully reread your original prompt and the associated materials (Ex06c and the playground from Ex09a).

→ Clearly specify the task ID and the name of the LLM that was used, the exact issues with the results, and the potential reasons behind the unsatisfactory performance

Write here

Task ID: [   ]   Type of task: [   ]   Model used: [   ]

What were the problems with the result you got?
Why did the LLM perform the way it did?
What went wrong?

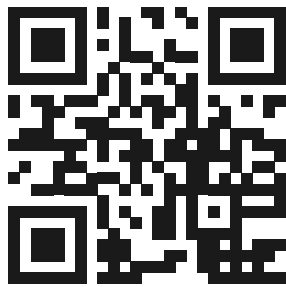# EX^10b - *Learning* Your Craft

Instructions

You will now explore various prompting techniques designed to improve interactions with LLMs. A prompting technique is a standardized way to write prompts that is supposed to improve the results we get out of LLMs.

We have curated a set of resources and tutorials for you to learn these methods.

Dedicate 60 to 90 minutes to reviewing these resources.
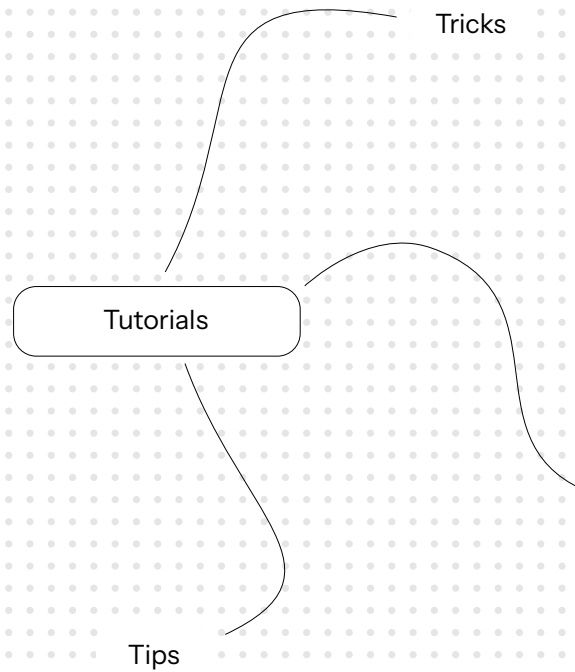
As you study each one, reflect on these techniques critically.

List of tutorials

Notice

Use the blank mind map on the next page to keep track of the tips and tricks you learned in the tutorials.

EX^{10b} – *Learning* Your Craft / 1

Keep track of tips and tricks:

Tricks

Tutorials

Tips

EX[10b] – *Learning* Your Craft / 2

Do these techniques apply to your original prompt?

After exploring these techniques, do you now feel differently about how much your original prompt shaped the results?

EX^{10c} – *Practice* Makes Perfect

Instructions

Using the prompting techniques you've learned, revise your original prompt. Conduct multiple trials. You can test various techniques individually and in combination. You could also create new chats using the same prompt each time to assess consistency.

Notice

Answer to the questions on the following page

EX^{10c} - *Practice Makes Perfect*

What specific adjustments did you make?

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

How did the LLM's performance change?

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

LLM   TECHNIQUE   YOU

LLM   TECHNIQUE   YOU

LLM   TECHNIQUE   YOU

LLM   TECHNIQUE   YOU

LLM   TECHNIQUE   YOU

LLM   TECHNIQUE   YOU

LLM   TECHNIQUE   YOU

LLM   TECHNIQUE   YOU

EX^{10}

LLM   TECHNIQUE   YOU

LLM   TECHNIQUE   YOU

LLM   TECHNIQUE   YOU

LLM   TECHNIQUE   YOU

LLM   TECHNIQUE   YOU

LLM   TECHNIQUE   YOU

EX^{10d} – Detective work / *Playground*

Instructions

Picture yourself as a detective uncovering the mysteries behind your revised prompts and their outcomes. Dive into your conversations from Exercise 10c and hunt down clues to identify patterns:

→ What explains why some prompts worked while others fell short?

Carefully examine the interactions between the prompt, the chosen technique, and the LLM — were they partners in crime or suspects in disguise? Do they have different motives or alibis?

Create your own «detective wall», a visual map of your findings and theories about what actually happened. Each box represents a possible fact, observation, witness, suspect or culprit in your investigation.

Each box can belong to three different families:

→ *The LLM.* For issues related to the machine (limited understanding, hallucination, generic answer, ...)

→ *The Technique.* For issues related to the prompting, whether the technique itself (efficacy, applicability, ...) or the tutorial (clearly explained, easy to apply, ...)

→ *Yourself.* For issues related to your own actions (whether you adapted the prompting technique successfully or not, or you provided insufficient or too much information, ...)

When you add an element to your detective wall, tick off a box (LLM, Technique, You). Draw connections between the elements to visualize the relationships, alliances and conflicts.

After your investigation, narrow it down to the true culprit(s) and write down what happened in the center of the playground.

Example

Here the model wasn't able to grasp what I was asking and confused the "mandarin" food with the language name

EX¹¹ – Tracking *Shifts*

Instructions

Reflect on your experiments with LLMs since you began the exercises. Browse through your vademecum and your conversation history in ChatGPT or other LLMs. Answer the following questions while keeping in mind these two principles:

→ *Going wide.* Try to think about all aspects of your practice: technical, professional, emotional, ethical, and so on.

→ *Documenting your practice.* You must provide specific examples to support your answers.

Notice

Answer to the questions on the next page.

EX[11] - *Tracking shifts*

How has your practice evolved?

EX[11] – *Tracking shifts*

What aspects of your practice have remained consistent?

Additional notes