

Appendix 1 – Prompts and applied patterns

Prompts and applied patterns to generate the blueprint

Our prompting strategy was to break down the overall prompt into smaller steps. We decided to break down a prompt into smaller steps each time ChatGPT started to invent or repeat output. In the end we found it necessary to combine two manual preparatory steps with five prompts for completing the subtasks. To refine the prompts, we used the Alternative Approaches pattern - letting ChatGPT suggest alternative prompts - and in case a prompt led to no results the Refusal Breaker Pattern - ChatGPT suggesting alternative prompt formulation it can conduct [1,2]. In-between results can be provided upon reasonable request to the corresponding author.

	Manual step or prompt	Patterns [1]	Comment / result
1	Manually obtain the list of disease groups of the MeSH category "Diseases" (C)	-	20 disease groups (excluded: C07, C11, C22, C23)
2	Prompt for each of the MeSH disease category: "Act as an epidemiologist and provide a downloadable CSV file of the top 20 [<i>MeSH disease C1-23</i>] and their incidence in the format cases/100.000/year for the population in Europe in 2023. At the end, please include the source you have obtained that data from."	Persona pattern, Fact check list	We ran prompts up to three times each and picked best match
3	Manually compile all 20 sublists into one list and remove duplicates. If incidences do not match for duplicates, take the average value. Upload disease file and the list of 40 key symptoms from the iCoViP project	-	resulted in a list of 301 diseases and their incidences (no manual corrections were made)
4	Prompt: "Please read these two documents. The first one contains a list of common key symptoms. The second is a list of approximate incidences per 100.000 / year of diseases in Europe. We will need these documents for the next steps."		
5	Prompt: "Based on these two files, we will create an outline of 200 virtual patient cases for medical students to develop clinical reasoning skills. We will do this step-by-step. Step 1. As a medical educator, manually compile a list of 200 diseases for the virtual patients from the file so that the overall collection is an authentic representation of the European patient population and roughly matches the given incidences but also includes a few of the rarer diseases. Please print this list including frequencies of each disease."	Audience Persona pattern, Persona pattern,	

6	Prompt: "Step 2: As a next step please create a table with each disease in multiple rows depending on their frequencies and combine each disease instance with one more or less typical key symptom from the list. The combination of disease and key symptom can also include some rarer onsets and should provide a variety of typical onsets, but it is crucial that the symptoms logically align with the diagnoses. Please create a table of all 200 disease-key symptom combinations. After you have created the table, please check again whether the key symptoms are aligned with the diagnosis and if not, please make adaptations. Finally, please provide the table as a downloadable file."	Fact check list pattern	
7	Prompt: "Step 3: Based on this table of diseases and key symptoms you have just created; we now manually generate an outline for 200 virtual patient cases for medical students to develop clinical reasoning skills. When creating the list, ensure an internal logic and consistency for each case outline. It is also very important to ensure that all demographic data reflect a realistic representation of Europe's patient population and consultation settings, based on validated European statistics. Each case should now in addition to the disease and key symptom include the following metadata: - age of the patient, including children and very old patients aligned with the age distribution of the European patient population and aligned with the disease - gender aligned with the gender distribution of the European population including trans female or trans male (if patient is at least a teen) and aligned with the disease- age of the patient, including children aligned with the age distribution of the European patient population and aligned with the disease After you have created the table, please go through it and check all new columns whether they really are representative of the European population. If not, please make necessary adaptations. Please create an updated table of all 200 case outlines."	Audience persona patterns, Fact check list pattern	
8	Prompt: "Step 4: Please add now to the table the following metadata:		

	- when age ≥ 16: sexual orientation, which is aligned with the distribution of the sexual orientation of the European population - profession, aligned with the profession distribution in Europe and considering the age of the patient. Consider that children under the age of 6 are not students and elderly over the age of 65 are probably retired. Include academic and non-academic jobs from all work-sectors and do not use retired for patients < 50. - when age ≥ 16: BMI aligned with the BMI distribution of the European population and aligned with the disease - ethnicity, aligned with the distribution of ethnicities in the European population - whether the patient has a disability (and its type), aligned with the distribution of disabilities in Europe - whether the patient has a migration background (and its origin), aligned with the distribution of migration background in the European population. - acuity of onset (acute, chronic, subacute) aligned with the key symptom and diagnosis. - setting of the consultation aligned with the acuity of onset and disease After you have created the expanded table, please go through it and check all new columns whether they really are representative of the European population and check the internal consistency of each case. If not, please make necessary adaptations. Please create a downloadable table of all 200 case outlines."		
9	Prompt: "Step 5: Now expand the table with a fictive first name and last name for each patient that is aligned with the gender and migration background. The name combinations need to be unique within this table and each first name and last name should not appear more than 3 times. After you have created the expanded table, please go through it and check again and ensure that each first name and each last name does not appear more than 3 times within the table. Please create a downloadable table of all 200 case outlines."		
10	Prompt: "please evaluate the created blueprint for each column whether it is representative of the patient population in Europe. Provide for each column a comment on whether it is representative or whether and which aspects are over- or underrepresented."	Fact check list pattern	see assessment below

11	Prompt: "Please provide an updated version of the table, fixing the issues you have detected for age, migration background, profession, and consultation settings."		Result: Final blueprint (see Appendix 1)
12	Prompt: "Manually check row-by-row the internal consistency of each case outline based on the defined steps. Make adaptations if necessary and provide a detailed log of the changes."		in our case did not improve the blueprint but got worse for the names.

[1] White J, Fu Q, Hays S, Sandborn M, Olea C, Gilbert H, et al. A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. arXiv; 2023. Available from:

<http://arxiv.org/abs/2302.11382>

[2] <https://platform.openai.com/docs/guides/prompt-engineering>

Prompts to generate the VPs

The following prompt generated the pilot VP, all other VPs were created with the same prompt by replacing the VP-specific data from the blueprint. The prompts are based on the same patterns as the blueprint prompt.

As a medical educator please create a virtual patient case for the following patient:

Name: James Gonzalez

Age 47-years old

Sexual orientation: heterosexual

Migration Background: no

Disability: none

Profession: Physician

Gender: male

BMI: 18.7

key symptom: subacute onset of cough

Setting: primary care

Final diagnosis: community acquired pneumonia.

When manually writing the case scenario, consider the following rules when creating the case:

- Case is written in a way that puts the reader into the role of the healthcare professional
- Case uses direct speech dialogs between the healthcare professional (= the reader), the patient and if applicable other persons
- Case should be patient-oriented
- Case considers a realistic timeline depending on the setting of the encounter and the availability of technical examinations. This might require multiple encounters and transfers.
- All of the above stated data about the patient should be recognizable in the case scenario, but, if possible, should be indirect (e.g. sexual orientation can be shown by a wife/husband accompanying a patient)
- Case is based on medical and clinical guidelines relevant for the symptoms and the patient.
- Case starts with the primary contact between the healthcare professional and the patient and the patient explaining the key symptom

- Case includes then all steps from first impression, medical history (in dialog form), full (not only symptom oriented) physical exam reporting all findings (not only pathological), technical examinations including imaging and laboratory findings when appropriate to the diagnosis and/or the treatment of the patient
 - Case includes a closing information about the long-term development (e.g. full recovery, mild or severe mid- or long-term aftereffects, death).
 - Case can include specialist consultations if relevant and appropriate
 - After the physical exam provide a 1-2 sentence summary statement summarizing the current findings for the patient using semantic qualifiers
 - Case includes all results from the history and examinations, if an image or video for an examination result, such as x-ray, CT or MRI is needed, put "IMAGE NEEDED" with a description at the appropriate position in the case.
 - if lab results are obtained for the patient, include realistic results of a typical full lab test in a table, not only the pathological ones. Include also the normal ranges for each parameter.
 - results of any technical examinations should not be directly interpreted in the scenario, but be interpreted by the students (e.g. by using a question)
 - Where didactically appropriate include a question with answer options, the correct answers, and feedback, for example to help students identify pathological findings in lab results or an x-ray. The questions can be of different types depending on the didactic purpose, e.g. Multiple or single choice, freetext, sorting, or mapping
 - structure the case into logical sections with headings
 - After all necessary data has been provided in the case, add the following to the case: "STUDENT NEED TO DECIDE ABOUT DIAGNOSIS NOW" do not reveal the diagnosis before that step. After that the story can continue with the physician explaining the diagnosis to the patient.
 - Close the case scenario with a selection of up to three references and/or guidelines appropriate for medical students and relevant for the case content. For example, <https://www.statpearls.com/> or <https://www.uptodate.com> might have useful pages to provide.
- After having created the case scenario, go through it and check whether you have considered all the rules. make adaptations where necessary.
- Please create the case as a HTML downloadable file.