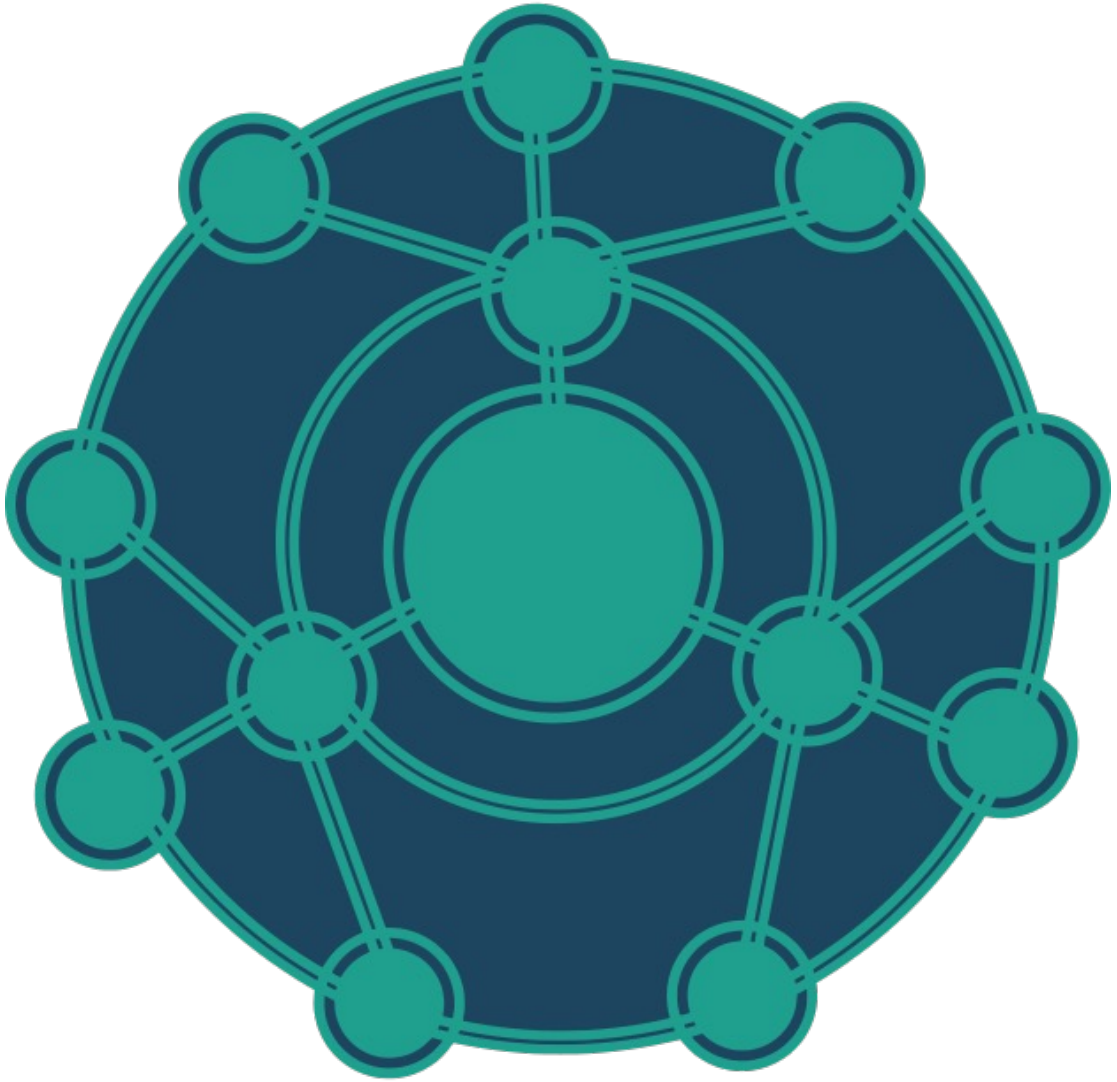
Estima

ISPOR Issue Panel

Embracing the AI Revolution: Exploring the Promising Future of Generative AI in HEOR

Julia Langham



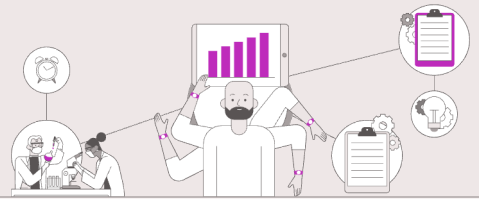
Agenda

01 Generative AI
- What's changed?

02 The HTA "Timeline"
- Where does generative AI fit in…?

03 Is Generative AI ready for HEOR?

Generative AI – What's changed?



As a company we were previously skeptical about the role of AI in HEOR…

Previous attempts at AI in HEOR:

- Were too specific for each purpose
- Required huge amounts of training data
- Still didn't work very well!

OpenAI have released a series of large language models (LLMs), most recently, GPT-4 (turbo)

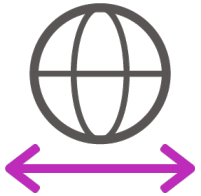
The major difference between LLMs and traditional AI models is (1) that they have been pre-trained on a massive amount of data and can be adapted to a wide range of tasks (be 'zero-shot' or finely tuned on new data) e.g. GPT-4 has been trained on 1 petabyte of data. (2) and can produce new data instances.

Now we have general purpose AI that can be deployed in HEOR with minimal effort

We are moving to a world where with the right software/processes HEOR projects will be done with natural language commands

With new more powerful models coming out (e.g., GPT-4 Turbo) and with things like "fine tuning" this will all get even better quickly!

The HTA “Timeline” - Where does generative AI fit in…?

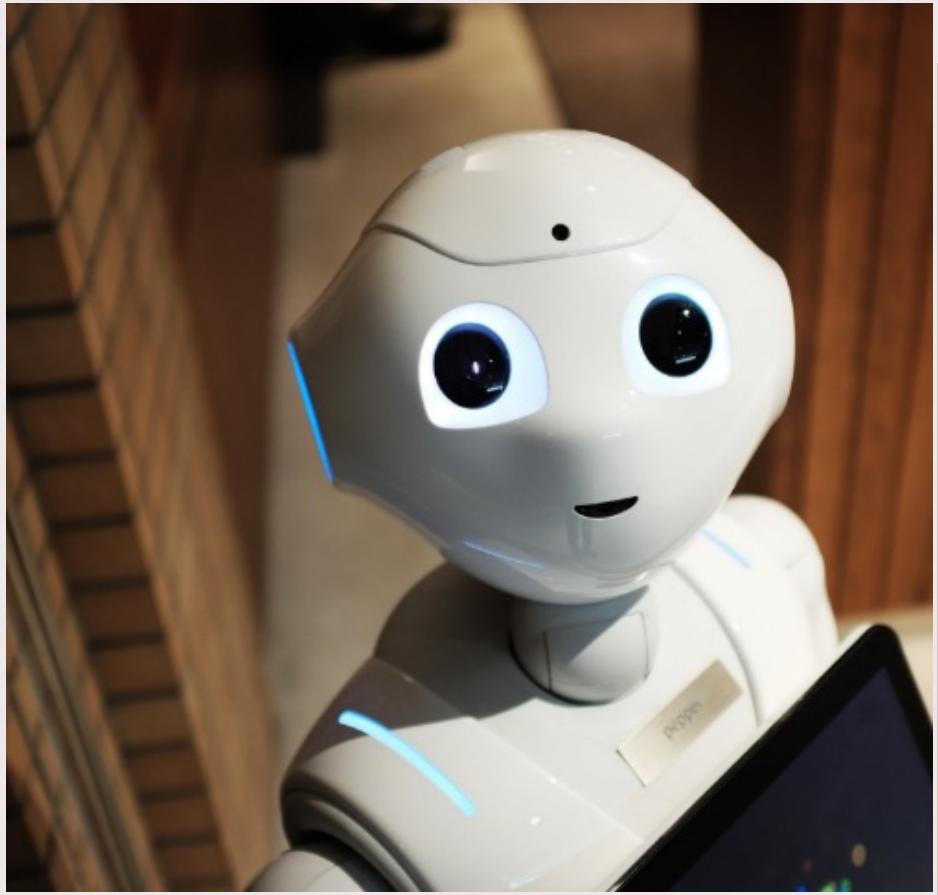


Systematic Literature Review (SLR)

Network Meta Analysis (NMA)

Economic Modelling

Everywhere!



End to end SLR

01 Generate search strings

02 Screen titles and abstracts, and full text

03 Apply QA checklists (Cochrane RoB)

04 Extract data

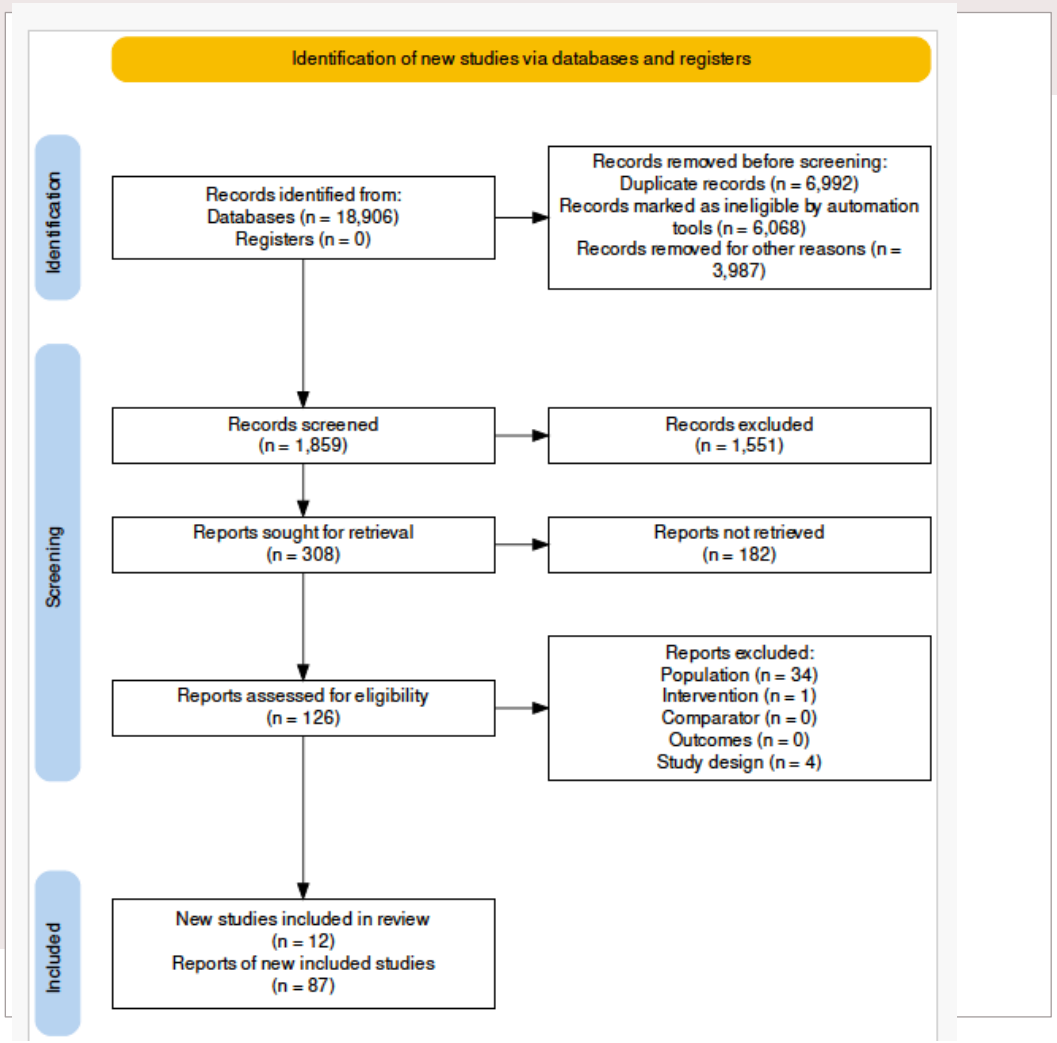
05 Reporting

Screening abstracts and full publications with GPT-4

- Tested on multiple examples including renal cell carcinoma, breast cancer and multiple myeloma
- Accuracy 85-90%; Sensitivity 95%-97%; Specificity 70%-90%
- The process can be incredibly fast – 10,000 abstracts in 10 minutes
- Very detailed data extraction sheets for quality assessment
- PRISMA flow diagram completion

Accurately identify eligible RCTs from titles & abstracts and from full publication reviews in a fraction of the time.

Highly adaptable to different reviews and ready for use

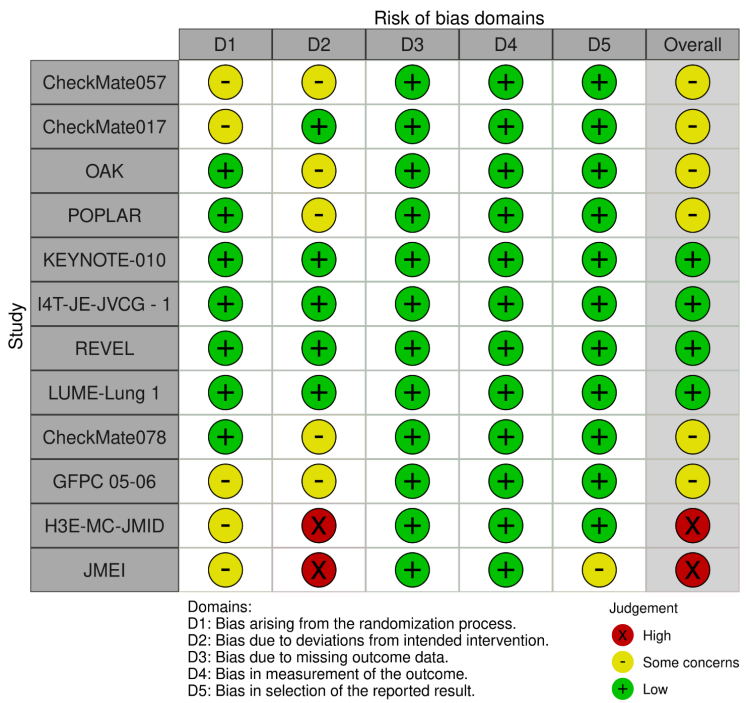


GPT-4 for RoB (2) assessment

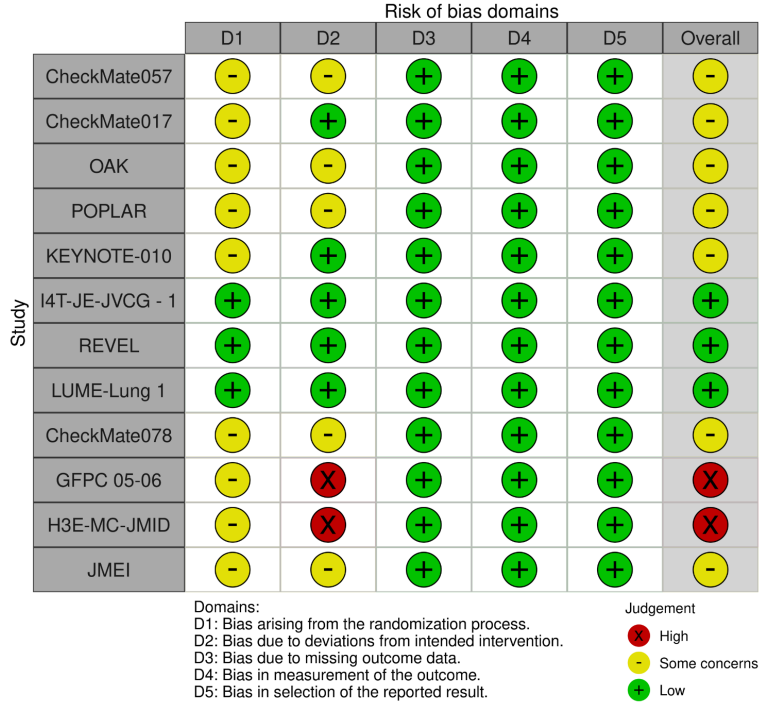
Very good agreement with human reviewers using the Cochrane risk of bias (revised) tool.

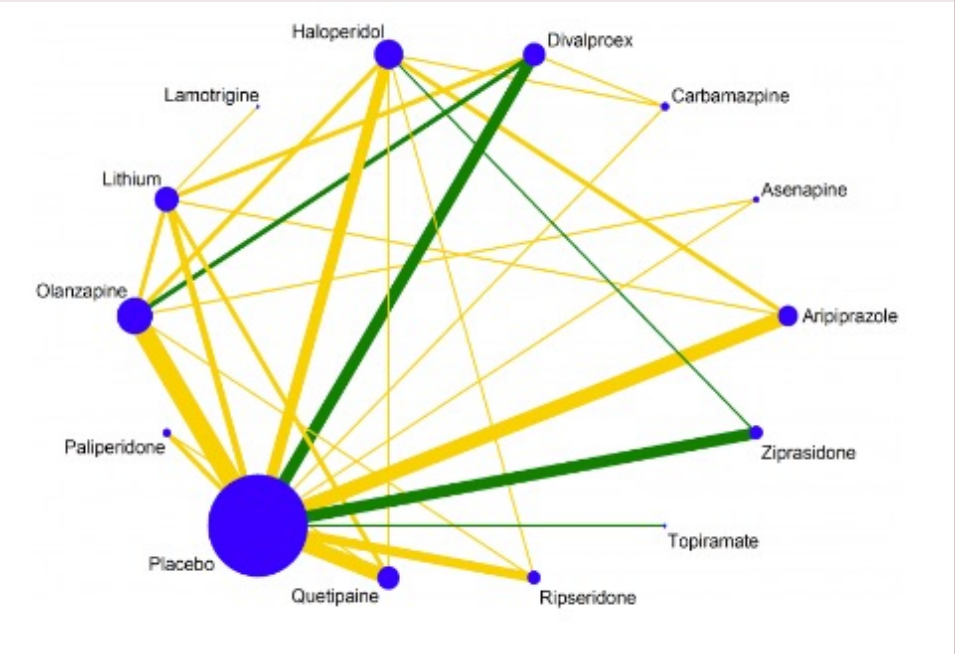
Ready to be tested and applied to new SLRs

GPT-4 RoB 2 Visualization



Human reviewer RoB 2 Visualization





NMA - What can Gen AI do?

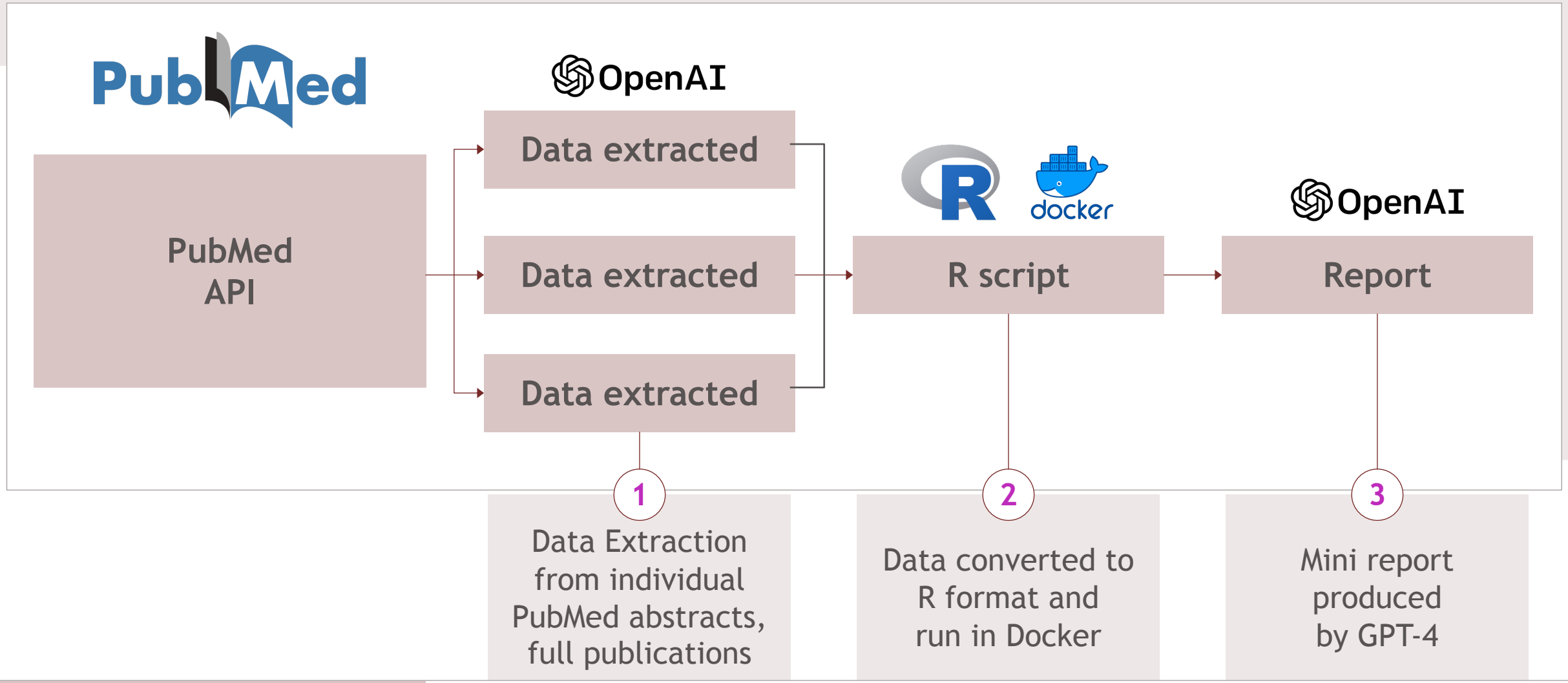
01 Conduct topline feasibility assessments

02 Extract data in NMA format

03 Run NMAs end to end

04 Write mini NMA reports

End to end NMA in GPT-4



GPT-4 for data extraction

Accurate and fast data extraction, in different formats

```
SUNSHINE Trial:
Treatment 1: Secukinumab every 2 weeks
Number of events 1: 81.5 (45%) of 181 patients had a hidradenitis suppurativa clinical response
Number at risk 1: 181
Treatment 2: Secukinumab every 4 weeks
Number of events 2: 75.2 (42%) of 180 patients had a hidradenitis suppurativa clinical response
Number at risk 2: 180
Treatment 3: Placebo
Number of events 3: 60.7 (34%) of 180 patients had a hidradenitis suppurativa clinical response
Number at risk 3: 180
Total trial size: 541
Randomisation ratio: 1:1:1

SUNRISE Trial:
Treatment 1: Secukinumab every 2 weeks
Number of events 1: 76.2 (42%) of 180 patients had a hidradenitis suppurativa clinical response
Number at risk 1: 180
Treatment 2: Secukinumab every 4 weeks
Number of events 2: 83.1 (46%) of 180 patients had a hidradenitis suppurativa clinical response
Number at risk 2: 180
Treatment 3: Placebo
Number of events 3: 57.1 (31%) of 183 patients had a hidradenitis suppurativa clinical response
Number at risk 3: 183
Total trial size: 543
Randomisation ratio: 1:1:1
```

NMA report produced by GPT-4 from end-to-end NMA

Correctly present and interpret results in a report

Introduction

Hidradenitis suppurativa (HS) is a chronic, inflammatory skin disease characterized by recurrent, painful nodules and abscesses, predominantly in the axillary, inguinal, and anogenital regions. The disease affects approximately 1-4% of the global population, with a higher prevalence in women and individuals of African descent. HS significantly impacts the quality of life of patients, leading to physical discomfort, psychological distress, and social isolation. Moreover, the disease poses a substantial economic burden on healthcare systems due to the high costs associated with long-term management and treatment.

Methods

We conducted a network meta-analysis (NMA) to compare the clinical response rates of different treatments for HS. The treatments included were placebo, ADA weekly, Secukinumab every 2 weeks, and Secukinumab every 4 weeks. The data were sourced from four clinical trials: PIONEER I, PIONEER II, SUNSHINE Trial, and SUNRISE Trial. The percentage of clinical responses was converted to the number of clinical responses, and a fixed-effects model was fitted to the data. The model was run on four chains with 2000 iterations each, and the first 1000 iterations of each chain were discarded as warm-up. The convergence of the model was assessed using the potential scale reduction factor (Rhat), where a value close to 1 indicates convergence.

Results

The model converged successfully, as indicated by the Rhat values of 1 for all parameters. The estimated odds ratios (ORs) for the treatments compared to placebo were as follows:

- ADA weekly: OR = 2.85, 95% CI = [2.02, 4.03]
- Secukinumab every 2 weeks: OR = 1.59, 95% CI = [1.19, 2.17]
- Secukinumab every 4 weeks: OR = 1.64, 95% CI = [1.20, 2.22]

Interpretation

The results suggest that all three treatments are more effective than placebo in achieving a clinical response in patients with HS. ADA weekly appears to be the most effective treatment, with an odds of achieving a clinical response approximately 2.85 times higher than placebo. The odds of achieving a clinical response with Secukinumab (either every 2 weeks or every 4 weeks) are approximately 1.6 times higher than placebo. However, the confidence intervals for the ORs overlap, indicating that the differences between the treatments may not be statistically significant. Further research is needed to confirm these findings and to investigate other factors that may influence the effectiveness of these treatments.

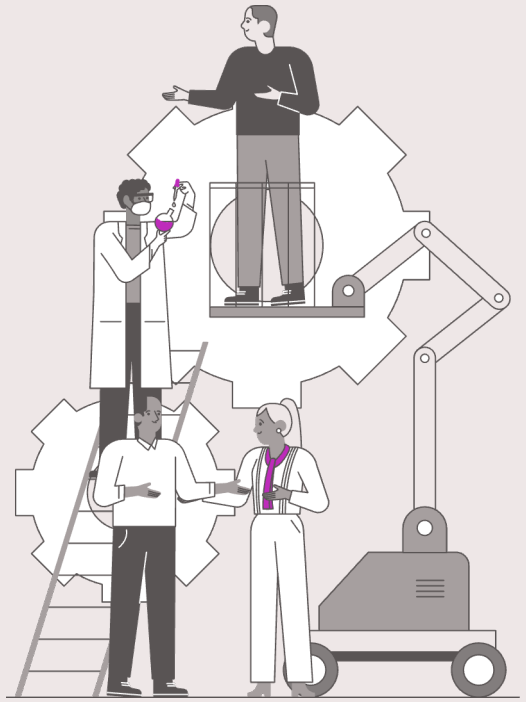
Modelling - What can GPT-4 do?

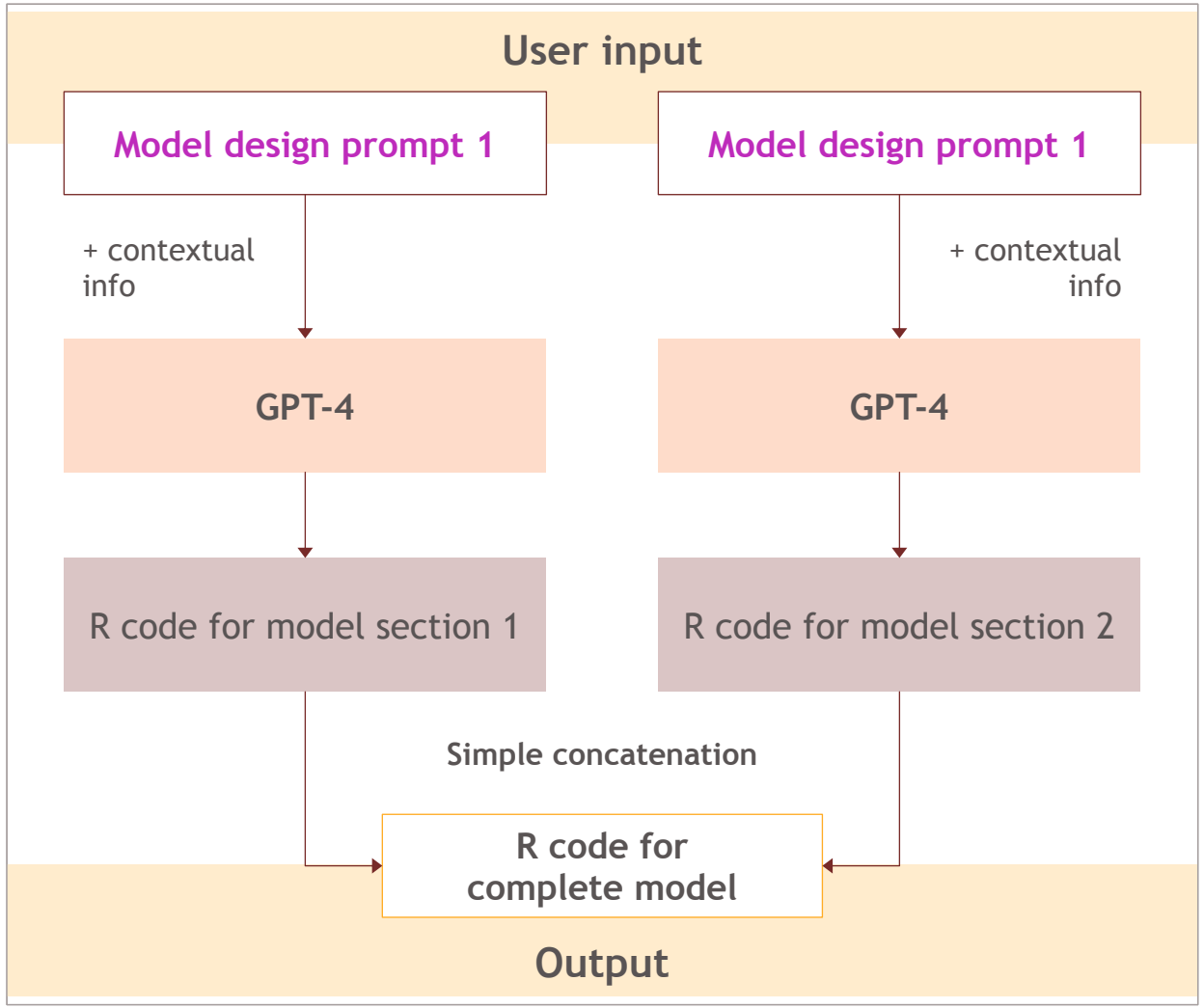
01 Replicate Health economic models in R

02 Basic Excel model adaptations

Model R script produced by GPT-4

Fully executable economic model adapted from shell script for a specific purpose using GPT-4 commands and our own process (below)



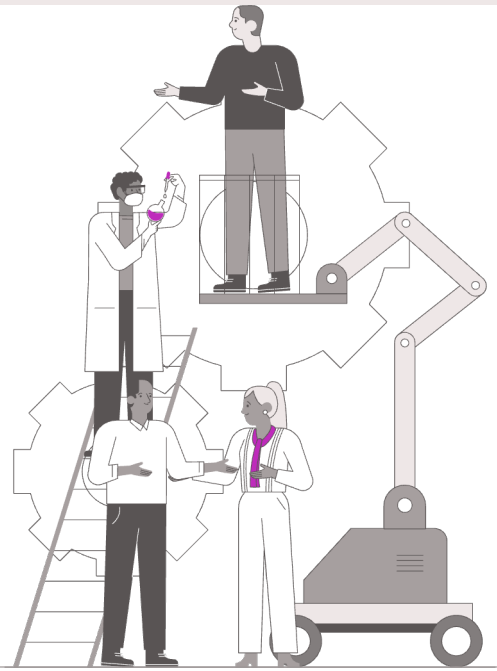


Model R script produced by GPT-4

Fully executable economic model adapted from shell script for a specific purpose using GPT-4 commands. Error-free models replicated the published ICERs to within 1%

```
###### Calculation: adverse event costs ######

### Nivolumab ###

# Step 1
n_cycles = time_horizon/cycle_length # assume time_horizon and cycle_length are defined elsewhere
nivolumab_ae_cost_array <- array(0, dim = c(n_cycles, 1, 3))

# Step 2
# The user has requested costs to be applied as a one-off cost in the first model cycle, so we will calculate the total cost of AEs for nivolumab and apply this in the first row of the total cost array

nivolumab_ae_cost <- sum(p_ae_freqs_nivolumab * p_ae_costs)

nivolumab_ae_cost_array[1,,] <- nivolumab_ae_cost

# Step 3.
nivolumab_ae_cost_array = nivolumab_ae_cost_array * nivolumab_dot_array

### Docetaxel ###

docetaxel_ae_cost_array <- array(0, dim = c(n_cycles, 1, 3))

# Step 2
# The user has requested costs to be applied as a one-off cost in the first model cycle, so we will calculate the total cost of AEs for docetaxel and apply this in the first row of the total cost array

docetaxel_ae_cost <- sum(p_ae_freqs_docetaxel * p_ae_costs)

docetaxel_ae_cost_array[1,,] <- docetaxel_ae_cost

# Step 3.
docetaxel_ae_cost_array = docetaxel_ae_cost_array * docetaxel_dot_array
```

GPT-4 ready for HEOR…



Successfully replicated the results of published systematic reviews, NMAs and economic models with a high degree of accuracy

Elements should start to be incorporated into the SLR and NMA workflow right now. A very powerful tool for improving workflow and efficiency

A very important requirement for human expertise and QA “human in the loop”

Early stage; AI models are improving fast, GPT-4 Turbo was released on November 6th which already removes some of the limitations encountered in these applications