PathoGeneMap

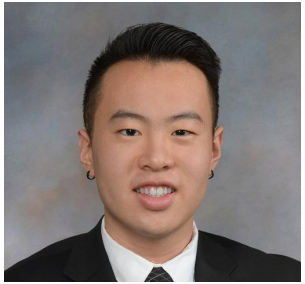
Empower biomedical researchers with AI-enhanced exploration of disease causing genes

Wesley Chang, Almicia Dunson,
Max Kaufmann, Christian Lee, Andrew McCall

OUR TEAM

Almicia Dunson

PM & Web Dev



Wesley Chang

Infrastructure & Web Dev

Max Kaufmann

SME & Data Engineering

Christian Lee

ML Engineering

Andrew McCall

Infrastructure & Data Engineering

THE PROBLEM

- Finding new treatment targets for human diseases
- More than 20,000 genes to choose from
- Which ones should I target?

THE PROBLEM

- Standing on the shoulders of giants

PubMed®

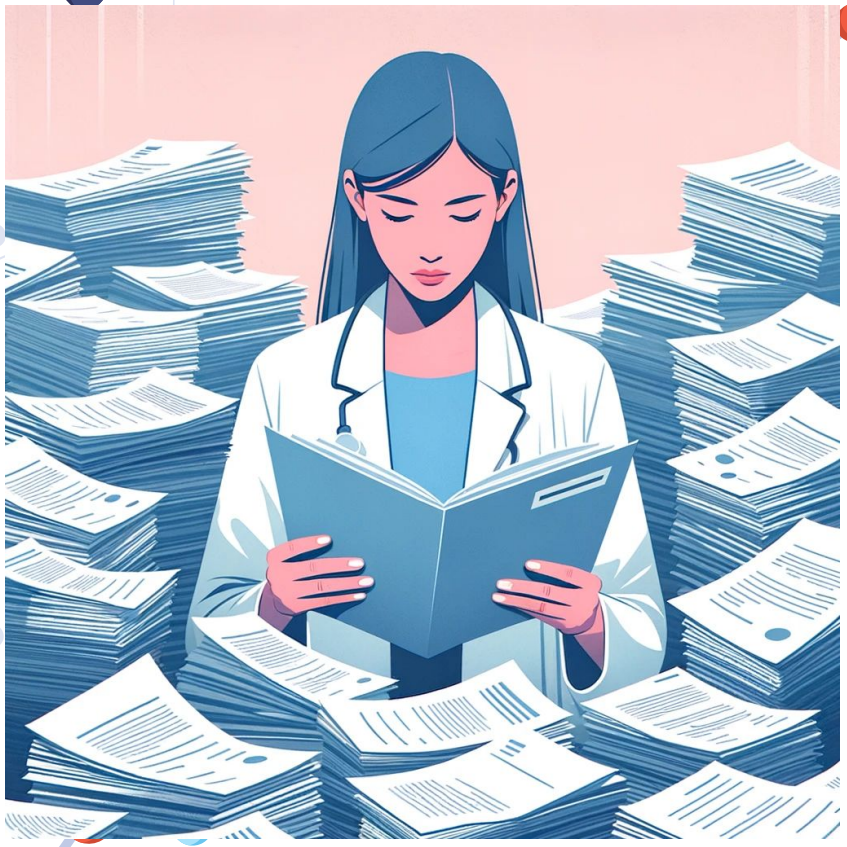
- Provides > 23 million biomedical literature abstracts

THE PROBLEM

Experiments in PubMed are described in **natural language only**!

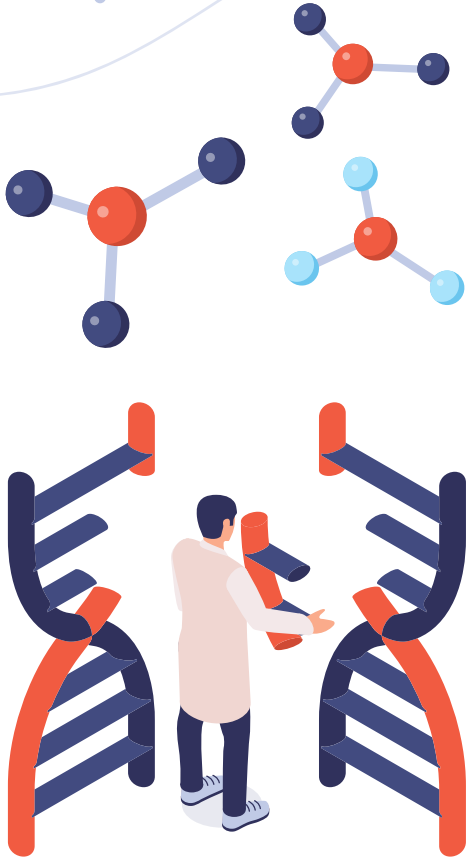
Intracerebral infection of mice with mouse hepatitis virus, a member of the Coronaviridae family, reproducibly results in an acute encephalomyelitis that progresses to a chronic demyelinating disease. The ensuing neuropathology during the chronic stage of disease is primarily immune mediated and similar to that of the human demyelinating disease multiple sclerosis. Secretion of chemokines within the CNS signals the infiltration of leukocytes, which results in destruction of white matter and neurological impairment. The CC chemokine ligand (CCL)5 is localized in white matter tracts undergoing demyelination, suggesting that this chemokine participates in the pathogenesis of disease by attracting inflammatory cells into the CNS. In this study, we administer a mAb directed against CCL5 to mice with established mouse hepatitis virus-induced demyelination and impaired motor skills. Anti-CCL5 treatment decreased T cell accumulation within the CNS based, in part, on viral Ag specificity, indicating the ability to differentially target select populations of T cells. In addition, administration of anti-CCL5 improved neurological function and significantly ($p < or = 0.005$) reduced the severity of demyelination and macrophage accumulation within the CNS. These results demonstrate that the severity of CNS disease can be reduced through the use of a neutralizing mAb directed against CCL5 in a viral model of demyelination.

Inhibition of CCL5 ameliorated a multiple sclerosis disease model



THE STATE OF THE ART IS …

… READING



OUR MISSION

Empower biomedical researchers with AI-enhanced exploration of disease causing genes.

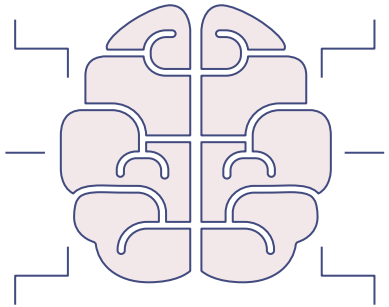
THE IMPACT

Number of genes interested in:	100
Amount of articles per gene:	900
Avg. minutes to determine article relevance:	2
Total days to find articles(8hr day):	375

A **couple of minutes** for a researcher to find relevant articles

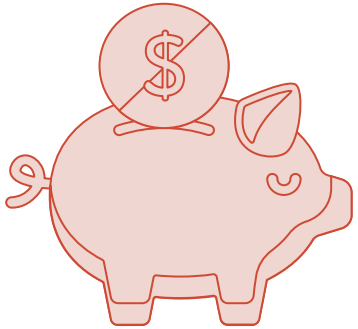
- **Time for a researcher without our product**
- **Time for a researcher *with* our product**

WHY HAS IT NOT BEEN DONE?



There is no model.

There is no labeled training data.



There is no funding.

THE TASK

Positive example

Negative example

Model input

Intracerebral infection of mice with mouse hepatitis virus, a member of the Coronaviridae family, reproducibly results in an acute encephalomyelitis that progresses to a chronic demyelinating disease. The ensuing neuropathology during the chronic stage of disease is primarily immune mediated and similar to that of the human demyelinating disease multiple sclerosis. Secretion of chemokines within the CNS signals the infiltration of leukocytes, which results in destruction of white matter and neurological impairment. The CC chemokine ligand (CCL)5 is localized in white matter tracts undergoing demyelination, suggesting that this chemokine participates in the pathogenesis of disease by attracting inflammatory cells into the CNS. In this study, we administer a mAb directed against CCL5 to mice with established mouse hepatitis virus-induced demyelination and impaired motor skills. Anti-CCL5 treatment decreased T cell accumulation within the CNS based, in part, on viral Ag specificity, indicating the ability to differentially target select populations of T cells. In addition, administration of anti-CCL5 improved neurological function and significantly ($p < or = 0.005$) reduced the severity of demyelination and macrophage accumulation within the CNS. These results demonstrate that the severity of CNS disease can be reduced through the use of a neutralizing mAb directed against CCL5 in a viral model of demyelination.

Experimental autoimmune encephalomyelitis (EAE) is a classical experimental model of multiple sclerosis (MS), an autoimmune disease of the central nervous system (CNS). Previous reports have suggested that matrine (MAT), a quinolizidine alkaloid derived from the herb Radix Sophorae Flave, could inhibit clinical EAE, but its mechanism of action is not clear. Our present study showed that MAT treatment resulted in dose-dependent reduction in neurological scores. Consistent with this observation, infiltration of inflammatory cells and demyelination in the CNS were also significantly suppressed. We further studied the mechanism underlying these effects of MAT by determining whether this treatment influences expression of molecules that are involved in the activation and migration of inflammatory cells. Our results showed that MAT significantly inhibited expression and production in the CNS of ICAM-1 and VCAM-1, key adhesive molecules, and CCL3 and CCL5, key chemokines, that attract inflammatory cells into the CNS. Furthermore, the TLR4/MD2 pathway, which plays an important role in the induction of Th1/Th17 cells in EAE, was also significantly inhibited. Together, our study not only demonstrates that MAT may be a novel therapeutic option for the treatment for MS, but also provides further information on mechanisms underlying the effect of MAT treatment.

Model output

Inhibition of CCL5 ameliorated a multiple sclerosis disease model

No specific gene was perturbed in vivo.

Database

Perturbation type	Inhibition
Gene	CCL5
Effect	Amelioration
Disease model	Multiple sclerosis

X

THE TASK

	Positive example	Negative example
Model input	Intracerebral infection of mice with mouse hepatitis virus, a member of the Coronaviridae family, reproducibly results in an acute encephalomyelitis that progresses to a chronic demyelinating disease. The ensuing neuropathology during the chronic stage of disease is primarily immune mediated and similar to that of the human demyelinating disease multiple sclerosis. Secretion of chemokines within the CNS signals the infiltration of leukocytes, which results in destruction of white matter and neurological impairment. The CC chemokine ligand (CCL)5 is localized in white matter tracts undergoing demyelination, suggesting that this chemokine participates in the pathogenesis of disease by attracting inflammatory cells into the CNS. In this study, we administer a mAb directed against CCL5 to mice with established mouse hepatitis virus-induced demyelination and impaired motor skills. Anti-CCL5 treatment decreased T cell accumulation within the CNS based, in part, on viral [illegible]ity, i[illegible]g the ability to [illegible]rentially tar[illegible] select popula[illegible] of T cells. In add[illegible]n, ad[illegible]stration [illegible] anti-CCL5 imp[illegible] neurolo[illegible] function and sig[illegible]y (p < or = 0.0[illegible]) red[illegible] the seve[illegible]f demyelinati[illegible] acr[illegible] accumulation wi[illegible]e CNS. These resul[illegible]onstrate tha[illegible] severity of [illegible] d[illegible] an reduced throu[illegible] he [illegible] of a neutrali[illegible] d[illegible]ainst CCL5 in [illegible] viral [illegible]odel demyelination.	Experimental autoimmune encephalomyelitis (EAE) is a classical experimental model of multiple sclerosis (MS), an autoimmune disease of the central nervous system (CNS). Previous reports have suggested that matrine (MAT), a quinolizidine alkaloid derived from the herb Radix Sophorae Flave, could inhibit clinical EAE, but its mechanism of action is not clear. Our present study showed that MAT treatment resulted in dose-dependent reduction in neurological scores. Consistent with this observation, infiltration of inflammatory cells and demyelination in the CNS were also significantly suppressed. We further studied the mechanism underlying these effects of MAT by determining whether this treatment influences expression of molecules that are involved in the activation and migration of inflammatory cells. Our results showed that MAT significantly [illegible]nhibited expr[illegible]on and prod[illegible]on in [illegible] ICAM-[illegible]nd VCAM[illegible] key adhesive [illegible]olecules, an[illegible]CL3 and CCL[illegible]ey ch[illegible]kines, th[illegible]ract [illegible]matory [illegible] into the CNS. [illegible]urthermore, [illegible] TLR4/MD2 [illegible]hich plays [illegible] imp[illegible]e in [illegible]e induction of Th1/Th17 cel[illegible]n EAE, was [illegible]so [illegible]ficantly in[illegible]ed. [illegible]study not only demonstrates [illegible] MAT may b[illegible] nove[illegible]apeutic [illegible]n fo[illegible] treatm[illegible] MS, but also [illegible]urth[illegible]n on [illegible]chanis[illegible] the eff[illegible]f MAT treatment.
Model output	Inhibition of CCL5 ameliorated a multiple sclerosis disease model	No specific gene was perturbed in vivo.
Database	(see table below)	X

Perturbation type	Inhibition
Gene	CCL5
Effect	Amelioration
Disease model	Multiple sclerosis

x 23 MILLION

HOW DOES GPT-4 PERFORM?

Test Data F1 Score

0.84

Cost to run on 23 Million Abstracts?

OUR MODELING APPROACH

PRE-TRAINED MODEL

Sci-Five

- T5-Large
- BioMedical Text
- Encoder-Decoder
- 770M parameters

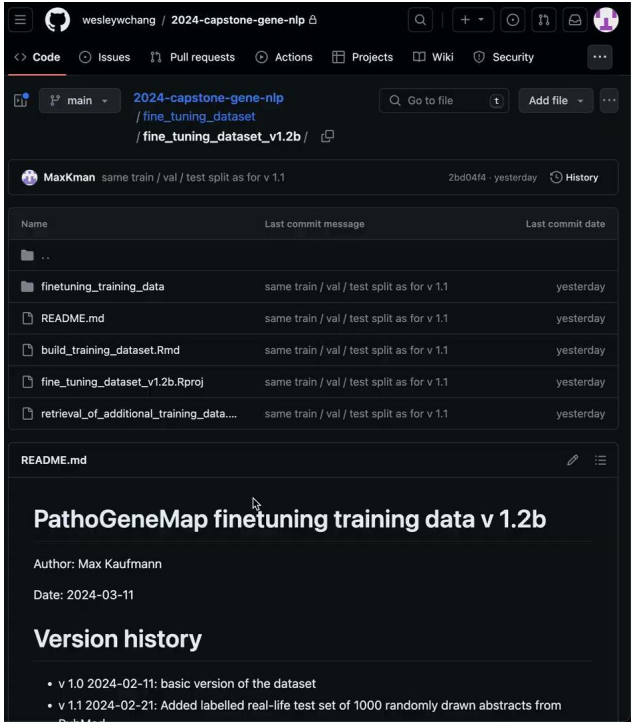
FINE TUNING

QLoRA

- Efficiency
- Faster Training Time
- Less GPUs needed

FINE TUNING TRAINING DATASET

- Own data set of **892** manually labelled abstracts for training
  - 50% positive examples +
    50% matched negative examples
- Public release for the benefit of the community



GPT4 VS OUR MODEL

Test Data F1 Score

0.84

Cost to run on 23 Million Abstracts

Test Data F1 Score

0.94

SCALING TO ALL OF PUBMED

Hardware

g5.xlarge

g4dn.xlarge

p3.2xlarge

.
.
.

•

Model

Quantization

LoRA

Merged Weights

.
.
.

•

Parameters

Input Token Length

Temperature

Batch Size

.
.
.

Minimize while maintaining performance

= { **Estimated Cost** }

SCALING TO ALL OF PUBMED

Hardware

Model

Parameters

Minimize while

g5

g4d

p3

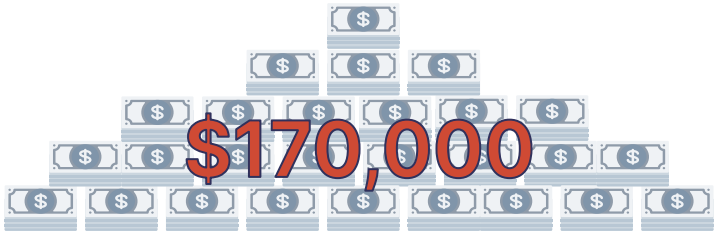
Completed inferences in parallel, reducing time from 41 days to 5 days

GPT4 VS OUR MODEL

Test Data F1 Score

0.84

Cost to run on 23 Million Abstracts?



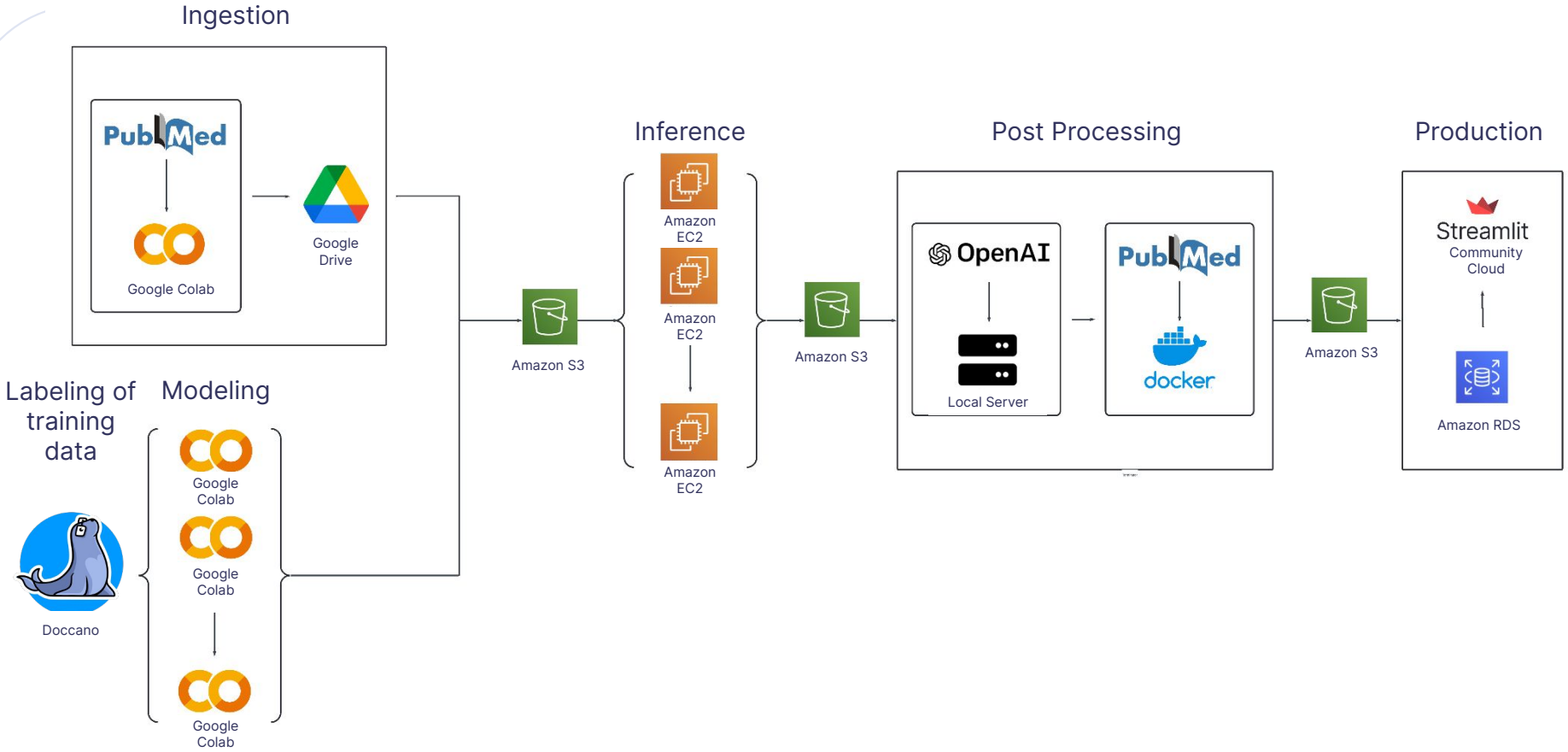
Test Data F1 Score

0.94

Cost to run on 23 Million Abstracts?

$820

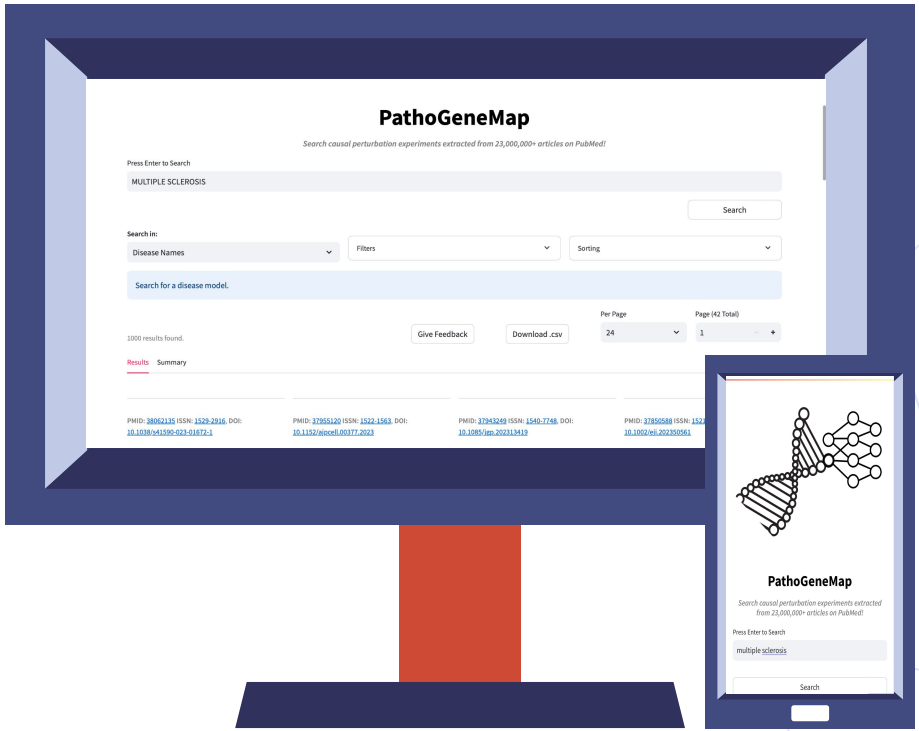
HOW DOES IT ALL FIT TOGETHER?



DEMO

WE ARE LIVE!

VISIT PATHOGENEMAP.COM AND TRY OUR APP



REACTIONS FROM OUR USERS

JAN-BRODER ENGLER

MD-PhD
Principal Investigator
University of Hamburg
Germany

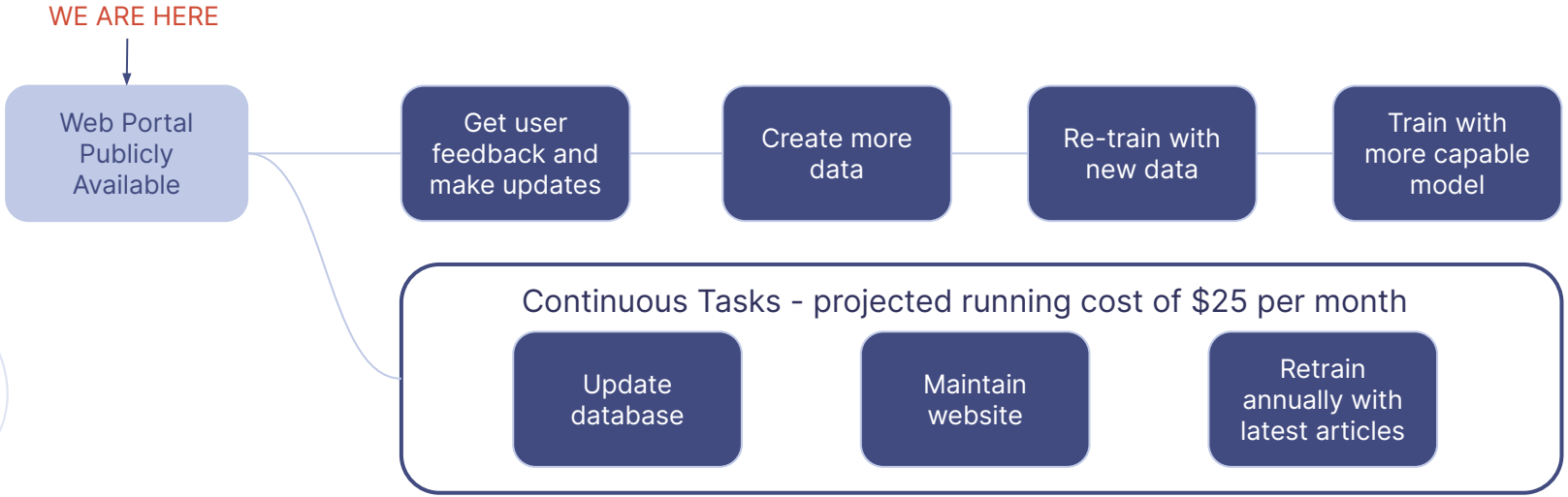
KATE ATTFIELD

PhD
Principal Investigator
University of Oxford
United Kingdom

"PathoGeneMap is an excellent starting point for conducting literature research on your gene of interest."

"A very easy to use tool, which is simple to navigate and gives immediate insight into potential gene functions that are placed directly in the relevant context"

PRODUCT ROADMAP



In Development

Planned for future

CONCLUSION

MISSION

Empower biomedical researchers with AI tools that help determine which genes cause diseases.

APPROACH

We fine tuned LLM's to read abstracts and extract gene information form them, then stored that information in a database that is accessible through a web portal.