



NEW COLOMBO PLAN

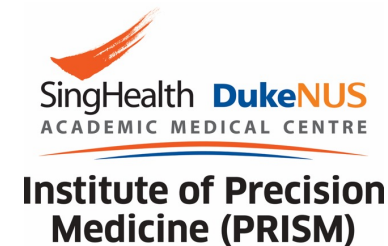
*Connect to Australia's future - study in the region*



Curtin University

# Clinical Free Text to HPO Codes

Gabrielle Stinton



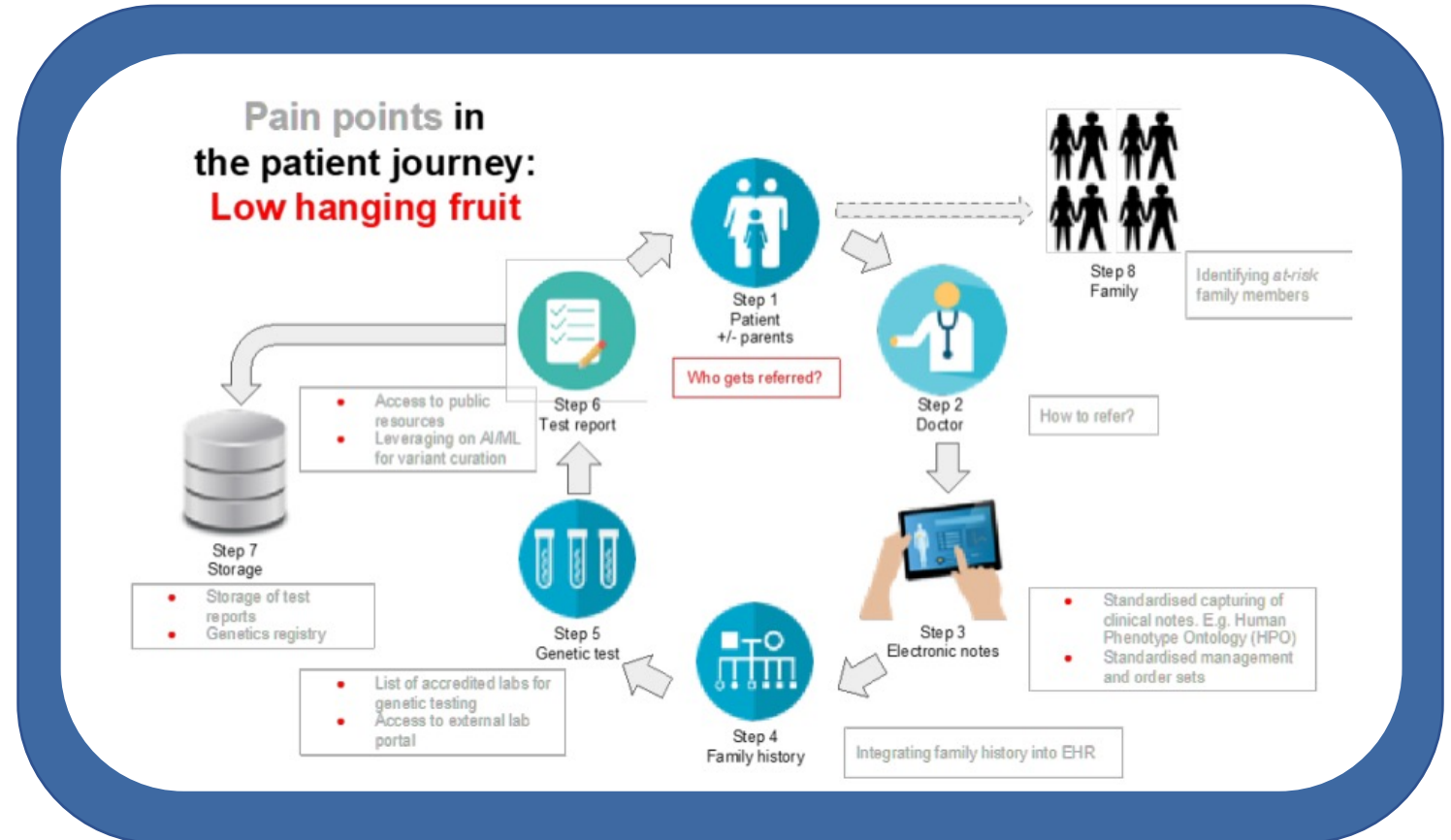
# Problem Statement

## Rare disease odyssey

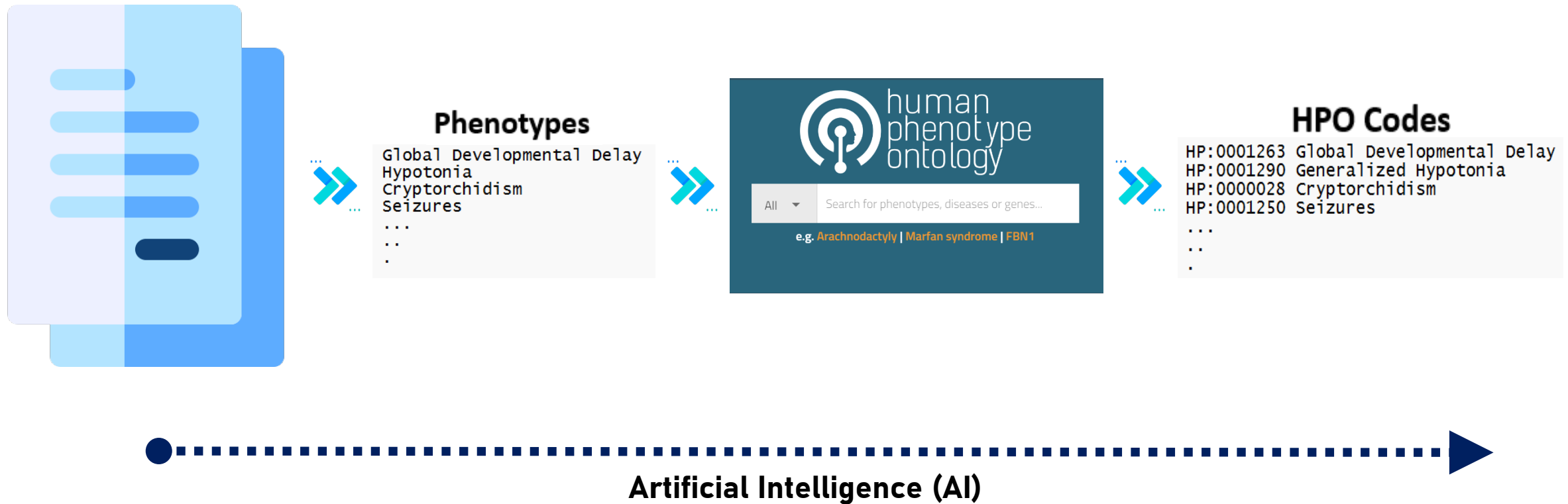
Decreasing diagnosis time,  
increasing diagnosis rates

## Clinical Notes

Standardized capturing of  
clinical notes into a genomic  
appropriate format e.g.,  
Human Phenotype Ontology  
(HPO)



# Current Methodology



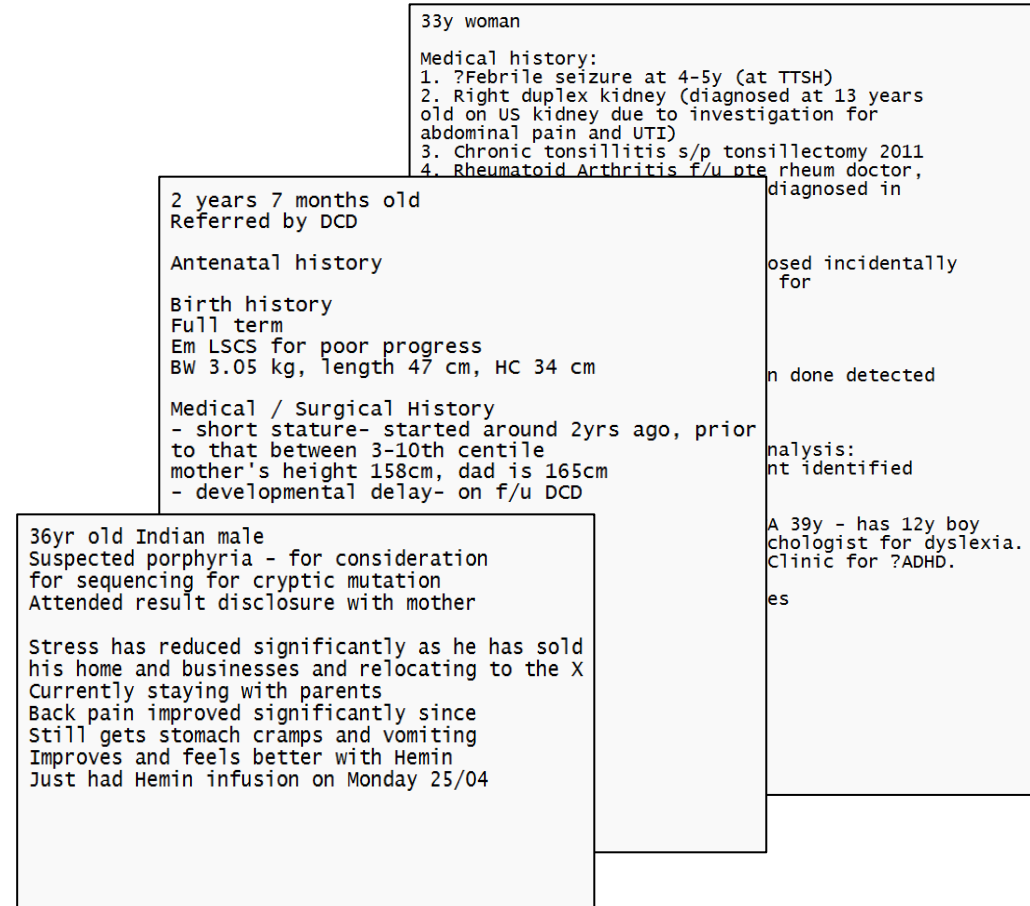
# Data

## Data Security

Access was organized by Dr Saumya, with reports de-identified before use

## Data Understanding

Average report length ~812 characters with phenotypes scattered in free text



# Challenges, Assumptions, and Limitations

---

## **Free Text:**

Synonyms, negation, abbreviations, overlapping concepts, numerical values, etc..

E.g., Hypotension (HP:0002615) can be expressed as “drop in blood pressure” and “BP of 79/48”

## **Sparse Multi-Class Classification Problem:**

There are > 13000 different HPO terms that need to be classified

## **Lack of a large public dataset:**

Need large amounts of manually HPO annotated clinical free text to train an ML/DL model

## **Limited data:**

Not able to accrue large amounts of data to train a model from scratch  $\Rightarrow$  pretrained models.



# Literature Review

Approach	Tools/Models	Semantic and Syntactic Meaning	Computational Efficiency
Dictionary-based	› National Centre for Biomedical Ontology (NCBO) annotator	✗	Very high
	› Monarch Initiative Program	✗	Very high
	› Cliniphen	✗	Very high
	› Open Biological and Biomedical Ontologies (OBO) annotator	✗	Very high
Machine Learning (ML)/ Deep Learning (DL)	› Neural Concept Recognizer (NCR) [1]	✓	Medium
	› Phenotagger [2]	✓	Low
	› PhenoBERT [3]	✓	High



# Ideation

---

## Key Considerations:

Medical use case means greater consideration on workflow integration

- › Handling model mistakes
- › Responsibility still with the clinician: human-in-the-loop

**Pre-trained model + human-in-the-loop**



# Web Application

A web app built in Streamlit that leverages the two current top performing models:  
PhenoBERT and Phenotagger





# Clinical Free Text to HPO Codes (Web App)

---

## Phenotagger Web Application

Based on Phenotagger web app (site is currently down)

### Customizations



- › comparison of annotation results with the current state-of-the-art (SOTA) model PhenoBERT



- › user selection of identified HPO codes and relevance categories



- › a search bar for users to enter phenotypes that the models may have missed which uses the Human Phenotype Ontology API [4]



# Example CDS Case – Manual

6 years old girl accompanied by mother and father

Attending EIPIC

Motor development:  
Still walking by herself, within 3 to 4 meter, stable

Speech development:  
- will shout for help to get attention.  
- will pull parents hand to point to things she wants through communication.

ARM rectovestibular - (3 stage repair because of IUGR and cardiac lesion)  
- Underwent sigmoid colostomy, PSARP and stoma closure  
2) Caudal regression/ spinal lipoma operated - 14/11/2011  
3) Neurogenic constipation on customised enema regimen  
4) Dysmorphism / term IUGR / global developmental delay  
- Attends EIPIC  
- CMA done - Normal  
- Planned for WES research  
5) Small muscular VSD, small ASD / PFO  
- last 2DE Sept 2013  
- Bedside 2DE 2017: Small anterior muscular VSD 3mm, left to right flow, unable to visualize ASD due to gaseous abdomen  
6) Bilateral dysplastic hips ( resolved ) and equinovarus feet ( AFOs)

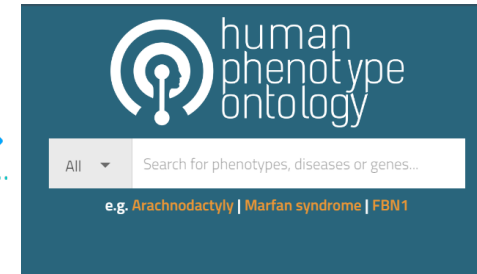
Examination:  
Alert, comfortable, active  
Heart S1S2  
Lungs clear  
Abdomen soft

congenital quint  
hypertelorism  
sacral dimple  
microcephaly  
nasal atresia  
imperforate anus (s/p repair)  
spinal cyst

Investigation done:  
CMA done - Normal

## Phenotypes

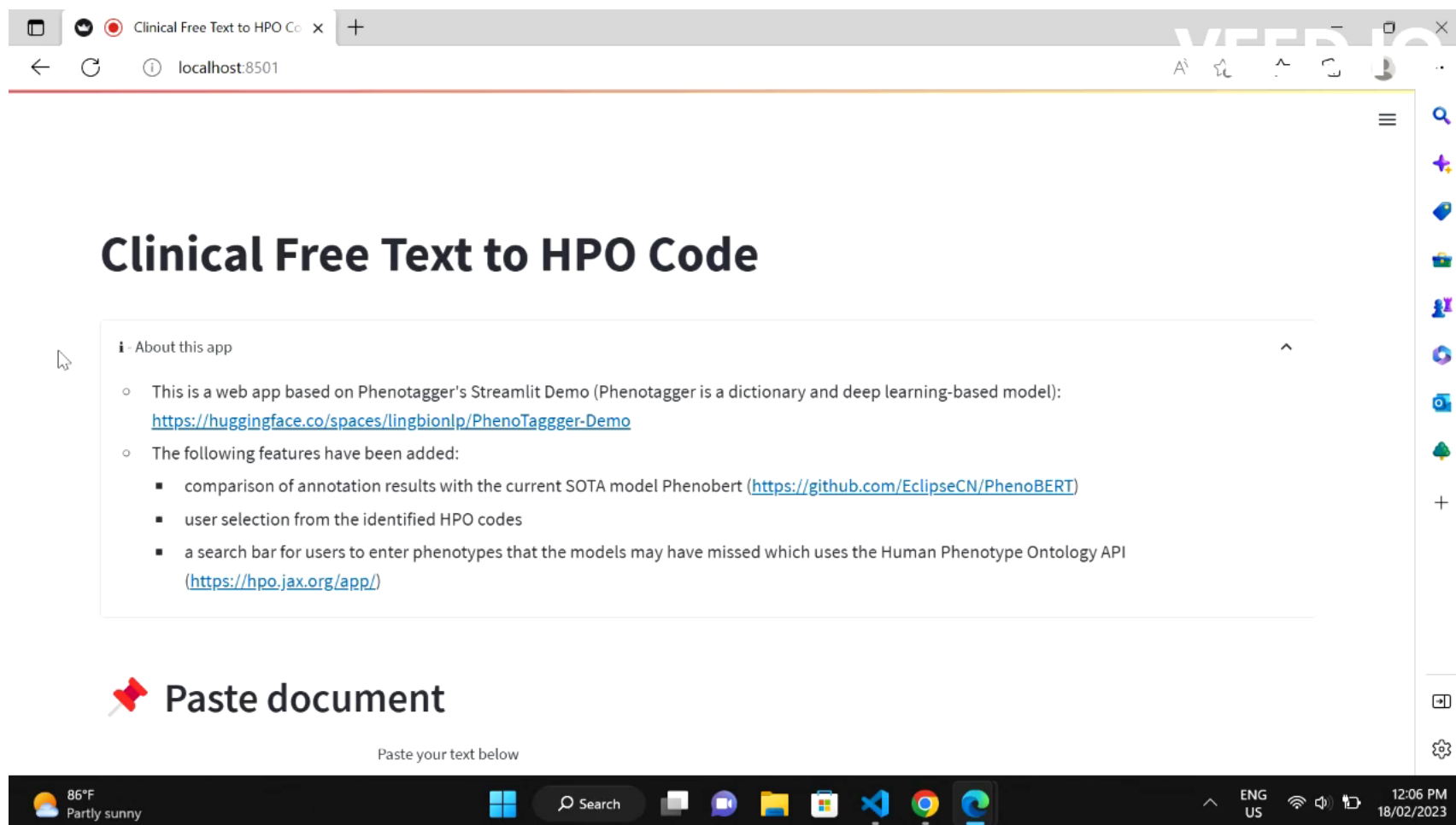
congenital squint  
hypertelorism  
sacral dimple  
microcephaly  
nasal atresia  
imperforate anus  
Spinal cyst  
Bilateral dysplastic Hip  
Bilateral equinovarus feet



## HPO Codes

HP:0000487 Congenital strabismus  
HP:0000316 Hypertelorism  
HP:0000960 Sacral dimple  
HP:0000252 Microcephaly  
HP:0010644 Midnasal stenosis  
HP:0002023 Anal atresia  
HP:0030726 Spinal neurenteric cyst  
HP:0001385 Hip dysplasia  
HP:0001762 Talipes equinovarus

# Example CDS Case – Web App



# Web App - Limitations

---

## Abbreviations

CDS52501

GDD - especially speech

## Negation

CDS56101

no scoliosis

## Misidentification

CDS52801

Plan to go back to Penang, Malaysia



kyphosis

## Detection Failures

CDS53801

Argininosuccinate Lyase Deficiency



HP:0025630 Argininosuccinic aciduria

# Case Study - ChatGPT

An advanced conversational chatbot developed by OpenAI and released on Nov 22 [5]



# Web App – Example Case 1

Telehealth on 9th June 2022 1pm ( both parents)  
Started off conversation wth father Mother joined at 1:07

According to father: Resting better at home Took alot of adjustments to get back to her routines

- Have been monitoring HR, improving, baseline now ~100bpm
- BP improved
- CIC minimal
- R hand strength improved
- Coordination improving
- Still needs some assistance with walking - still wobbly

Have not seen psychiatrist Parents keen on cutting back some medications

Discussed about mitochondrial disorders - for this particular condition, a higher percentage of heteroplasmy is associated with a probability of being affected However the percentage is not a reflection of disease severity

To continue cardiac monitoring To continue monitoring eyes - possibility of developing retinitis pigmentosa For respi followup, spirometry planned with Dr X

Father enquired if monitoring is for intervention

- for cardiac monitoring, there are medications that can be helpful, if required
- for eye, more for monitoring the development of RP. at this time, there are no drug therapies

Menarche April 2021 Weakness noted Dec 2021 Ptosis noted Jan 2022

X is likely considered within the mild to moderate range

Phenobert Results

HPO_id	Given Term	Official Term
<input type="checkbox"/> HP:0000510	developing retinitis pigmentosa	rod-cone dystrophy
<input type="checkbox"/> HP:0000508	ptosis	ptosis

Phenotagger Results

HPO_id	Given Term	Official Term
<input type="checkbox"/> HP:0000510	retinitis pigmentosa	rod - cone dystrophy
<input type="checkbox"/> HP:0025406	Weakness	asthenia
<input type="checkbox"/> HP:0000508	Ptosis	ptosis



# ChatGPT – Example Case 1

**G** annotate hpo codes from phenotypes and indicate what text is associated with each hpo code:

Telehealth on 9th June 2022 1pm ( both parents)

Started off conversation wth father  
Mother joined at 1:07

According to father:  
Resting better at home  
Took alot of adjustments to get back to her routines

- Have been monitoring HR, improving, baseline now ~100bpm
- BP improved
- CIC minimal
- R hand strength improved
- Coordination improving
- Still needs some assistance with walking - still wobbly

Have not seen psychiatrist  
Parents keen on cutting back some medications

Discussed about mitochondchondrial disorders - for this particular condition, a higher percentage of heteroplasmy is associated with a probability of being affected  
However the percentage is not a reflection of disease severity

To continue cardiac monitoring  
To continue monitoring eyes - possibility of developing retinitis pigmentosa  
For respi followup, spirometry planned with Dr

Father enquired if monitoring is for intervention  
- for cardiac monitoring, there are medications that can be helpful, if required  
- for eye, more for monitoring the development of RP. at this time, there are no drug therapies

Menarche April 2021  
Weakness noted Dec 2021  
Ptosis noted Jan 2022

X is likely considered within the mild to moderate range



- HP:0000855 Abnormality of blood pressure - "BP improved"
- HP:0003324 Cerebellar ataxia - "Still needs some assistance with walking - still wobbly"
- HP:0001290 Hypertonia - "R hand strength improved"
- HP:0000962 Ptosis - "Ptosis noted Jan 2022"
- HP:0003011 Weakness - "Weakness noted Dec 2021"
- HP:0012637 Abnormality of cardiac function - "To continue cardiac monitoring"
- HP:0000486 Visual impairment - "To continue monitoring eyes - possibility of developing retinitis pigmentosa"
- HP:0011017 Mitochondrial inheritance - "Discussed about mitochondrial disorders - for this particular condition, a higher percentage of heteroplasmy is associated with a probability of being affected."
- HP:0001607 Respiratory insufficiency - "For respi followup, spirometry planned with Dr"



# ChatGPT – Example Case 2

G

annotate hpo codes from phenotypes and indicate what text is associated with each hpo code:

5 years old boy

Has been well

Attending EIPIC in the afternoons, preschool in the morning - coping well  
Parent reports that there is improvement but feels that he is not ready for P1  
X is on follow up at NUH CDU

On follow up with KKH Cardiology for turbulence in branch PAs  
On follow up with KKH ENT for bilateral OME and ?high freq HL but mild

Development:

Climbs stairs, with support

Able to jump and hop

Recognises A to Z

Knows color

Speak in short sentences

Able to feed self

Toilet training - still wear diaper at school

O/E:

alert, comfortable, active

brachycephaly

no ridging of sutures

Heart S1S2 ESM

Lungs clear

Abdomen soft

1.



- HPO code: HP:0000252  
Text: Parent reports that X is not ready for P1
- HPO code: HP:0001635  
Text: Follow-up with KKH Cardiology for turbulence in branch PAs
- HPO code: HP:0000380  
Text: Bilateral OME (Otitis Media with Effusion)
- HPO code: HP:0002618  
Text: Mild high-frequency hearing loss
- HPO code: HP:0003124  
Text: Climbs stairs with support
- HPO code: HP:0001638  
Text: Able to jump and hop
- HPO code: HP:0002011  
Text: Able to feed self
- HPO code: HP: toilet training  
Text: Still wears diaper at school

2.



HPO codes:

- Brachycephaly (HP:0000248)
- Abnormal heart sound (HP:0001622)
- Clear lungs (HP:0012827)
- Soft abdomen (HP:0002027)
- Delayed toilet training (HP:0000746)
- Speech delay (HP:0000750)
- Abnormality of color vision (HP:0000551)
- Abnormality of the fontanelles or cranial sutures (HP:0000235)
- Difficulty climbing stairs (HP:0003551)
- Difficulty jumping (HP:0001967)
- Bilateral otitis media (HP:0000388)
- High frequency hearing loss (HP:0005101)





# Future Work/Recommendations

---

## Web App



Fine tuning



Dictionary for abbreviations



Negation



Future models

## ChatGPT



Limitations



Future versions



# Summary

---



The automatic conversion of clinical free text to HPO codes is just a small use case on how AI can be leveraged in the medical domain.

# Acknowledgements

---

- › Dr Saumya Shekhar Jamuar
- › Dr Sonny Pham
- › Dr Lim Weng Khong
- › Ms Sylvia Kam
- › Ms Lim Jiin Ying
- › Ms Jasmine Goh
- › Dr Sonia Davila
- › Ms Yasmin Byslra
- › Ms Simone Ng
- › CIRB 2019/2243





# NEW COLOMBO PLAN

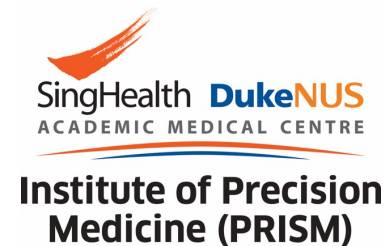
*Connect to Australia's future - study in the region*



## Curtin University

# Thank you!

Gabrielle Stinton



# References

---

- [1] A. Arbabi, D. R. Adams, S. Fidler, and M. Brudno, “Identifying clinical terms in medical text using ontology-guided machine learning,” *JMIR Medical Informatics*, vol. 7, no. 2, 2019.
- [2] L. Luo, S. Yan, P.-T. Lai, D. Veltri, A. Oler, S. Xirasagar, R. Ghosh, M. Similuk, P. N. Robinson, and Z. Lu, “Phenotagger: A hybrid method for phenotype concept recognition using human phenotype ontology,” *Bioinformatics*, vol. 37, no. 13, pp. 1884–1890, 2021.
- [3] Y. Feng, L. Qi, and W. Tian, “Phenobert: A combined deep learning method for automated recognition of human phenotype ontology,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, pp. 1–1, 2022.
- [4] Human phenotype ontology. [Online]. Available: <https://hpo.jax.org/app/>. [Accessed: 01-Feb-2023].
- [5] “CHATGPT: Optimizing language models for dialogue,” OpenAI, 02-Feb-2023. [Online]. Available: <https://openai.com/blog/chatgpt/>. [Accessed: 14-Feb-2023].

