



GOVERNMENT OUTCOMES LAB

2024 Spring Hack and Learn

21 March - Show and tell session



@golaboxford



golab.ox.ac.uk

Our agenda for today



GOVERNMENT
OUTCOMES
LAB

- Welcome and introductions
- Summary of our Hack and Learn event
- Challenges outputs, learnings and ideas for the future.
- Feedback from expert
- Q&A
- Social event!

The session is being recorded.

Please post questions in the chat throughout.

Hack and Learn phases



1. Kick off session: 07 March 2024

HACK HACK HACK -> use your Slack channels.

2. Show and tell session: 21 March 2024

Come and show your results and share your learnings. There will be a virtual social gathering after this session, all welcome!

Our Slack channels

The image displays three overlapping screenshots of a Slack workspace for 'GO Lab, University of Oxford'. The top-left screenshot shows a channel list on the left and a message in the '# hackteam33-chatgpt-and-data-extraction' channel from Shirley Haruka Bekins (3:59 PM, Thursday, March 14th) about re-reading instructions and using spreadsheet data. The middle-left screenshot shows a channel list with '# hack-and-learn-33-group-1' selected, and a message from ChatGPT (Friday, March 15th) about comparing report variables and using ChatGPT. The bottom-right screenshot shows a channel list with '# hack-and-learn-33-group-2' selected, and a message from ChatGPT (Tuesday, March 12th) about assisting with data extraction, including a file upload and replies from Juliana Outes and Petra Veres.

Our challenge



Data extraction for the Impact Bond Dataset using Artificial Intelligence -> [hackteam#33-chatgpt-and-data-extraction](#)



GOVERNMENT OUTCOMES LAB

Challenge#33!

Data extraction
using Artificial Intelligence - Group 1



@golaboxford



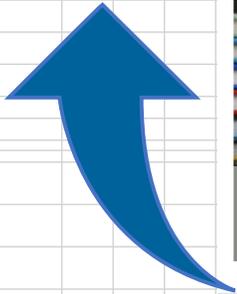
golab.ox.ac.uk

Task 1: Data Harvest Fest

Experimenting a data extraction pilot to test if chat GPT could read-up long policy reports and auto-complete the gigantic INDIGO impact bond spreadsheet.



Stage of Development	Value	Sources (IDs, comma separated)
	Complete	
Social Investment Prototype	Value	
Jurisdiction	Value	
Sources (IDs, comma separated)	Value	
Notes	Value	
Date outcomes contract signed	Value	
Date contracts between all parties signed	Value	
Anticipated completion date	Value	
Actual completion date	Value	
Start date of service provision	Value	
	2013-04	
Anticipated end date of service provision	Value	
Actual end date of service provision	Value	
Anticipated duration of service provision (in months)	Value	
Actual duration of service provision (in months)	Value	
Sources (IDs, comma separated)	Value	
	source5	
Notes	Value	
Overall project finance	Value	
Total investment commitment (currency)	Value	
	(GBP)	
	(min amount)	
	(exact amount)	3130000
	(max amount)	
	(min in USD)	
	(exact in USD)	4893220.107
	(max in USD)	
Maximum potential outcome payment (currency)	Value	

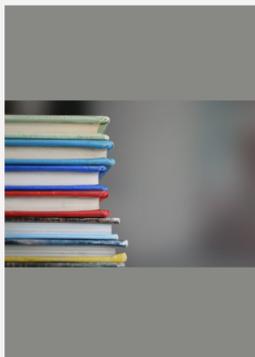


Evaluation Report | 6 Nov 2014

Evaluation of the Essex Multi- Systemic Therapy Social Impact Bond: Interim evaluation report

Organisation: Office for Public Management (OPM)

[More >](#)



Evaluation Report | 1 Oct 2016

Evaluation of the Essex Multi-Systemic Therapy social impact bond: findings from the first three years

Organisation: Office for Public Management (OPM)

[More >](#)

Evaluation Report | 6 Nov 2014

Evaluation of the Essex Multi- Systemic Therapy Social Impact Bond: Interim evaluation report

Organisation: Office for Public Management (OPM)

Evaluation Report | 1 Oct 2016

Evaluation of the Essex Multi-Systemic impact bond: findings from the first three months

Organisation: Office for Public Management (OPM)



GOVERNMENT OUTCOMES LAB



Evaluation Report | 7 Jul 2017

HMP Peterborough Social Impact Bond - cohort 2 and final cohort impact evaluation

Organisation: National Institute of Economic and Social Research [More >](#)



Evaluation Report | 9 Sep 2015

The payment by results Social Impact Bond pilot at HMP Peterborough: final project evaluation

Organisation: RAND Europe



Evaluation Report | 1 Dec 2022

Quality Education Development Impact Bond: A case study produced as part of the FCDO DIBs pilot evaluation

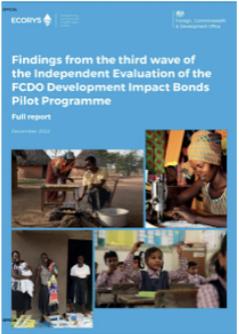
Organisations: Ecorys, Foreign, Commonwealth & Development Office (FCDO) [More >](#)



Evaluation Report | 4 Nov 2011

Lessons learned from the pilot implementation of the Social Impact Bond at HMP Peterborough

Organisation: RAND Europe



Evaluation Report | 1 Dec 2022

Findings from the third wave of the Independent Evaluation of the FCDO Development Impact Bonds Pilot Programme - Full Report

Organisations: Ecorys, Foreign, Commonwealth & Development Office (FCDO) [More >](#)

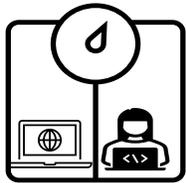
Task 2: Precision Peak Marathon

Investigating the accuracy and completeness of the data extracted using chatGPT



 The data extraction worked (well!!, somewhat) but how well, could it get it all? 

Does AI (chatGPT) perform better at data extraction when compared to data enthusiasts?



Did it modulate the tone while capturing the descriptive data?



Our learnings

Data Extraction using Chat GPT (Task 1)

- Chat GPT could speedily retrieve certain details with accuracy, specially, the quantitative variables.
- Chat GPT is not specifically attuned to the intricate details of INDIGO's outcome-based contracting data.
- Pre-emptive contextualization and strategic prompt generation foster robust results in data extraction.
- Chat GPT, currently struggles at handling segmented information and has limited tokens making data extraction challenging for data heavy projects.
- Strategic input and active verification in AI-assisted tasks, while also acknowledging the AI's limitations are the key to making the most of chat GPT.

Data accuracy and completeness (Task 2)

- Issues related to source access and data completeness, raising valid concerns about data credibility.
- The AI's analysis was informative, pinpointing critical gaps in detailed project data and outcome metrics.
- When extracting descriptive information, chat GPT was able to retain a fairly neutral tone, thus eliminating any bias.
- Tone and sentiment analysis on the report paved the way for a possible method to validate the AI-generated qualitative data's robustness and accuracy.



GOVERNMENT OUTCOMES LAB

Feedback from Expert



@golaboxford



golab.ox.ac.uk



GOVERNMENT OUTCOMES LAB

Challenge#33!

Data extraction
using Artificial Intelligence - Group 2



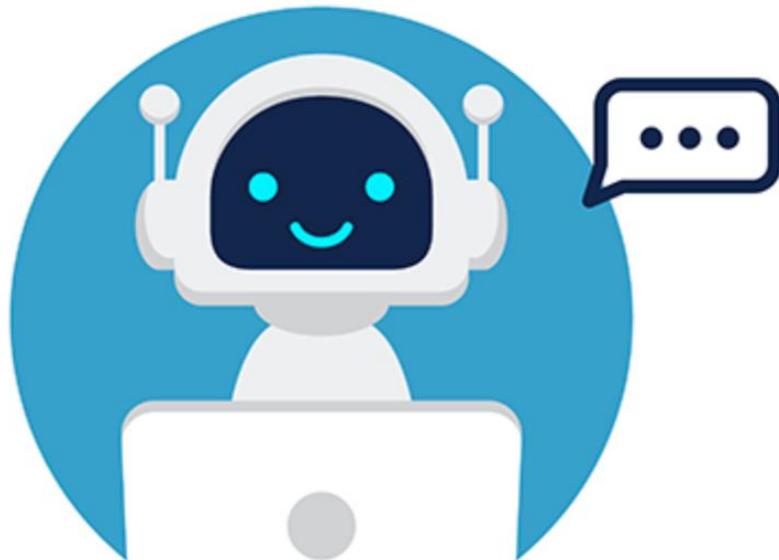
@golaboxford



golab.ox.ac.uk

Task 3: Auto Flow Fiesta

Creating and training a Chat-GPT bot to answer questions about impact bond projects



Create **Configure**



Name

Name your GPT

Description

Add a short description about what this GPT does

Instructions

What does this GPT do? How does it behave? What should it avoid doing?

Conversation starters

X

Knowledge

If you upload files under Knowledge, conversations with your GPT may include file contents. Files can be downloaded when Code Interpreter is enabled

Upload files

Preview



Start by defining your GPT.



Create **Configure**



Name
The Peterborough SIB

Description
This bot answers questions about the first social impact bond in the UK. Launched in 2010, the project aims

Instructions
GPT reads the documents and evaluation reports about the Peterborough SIB and uses that information to answer questions about the stakeholders and the project.

Conversation starters

Knowledge
If you upload files under Knowledge, conversations with your GPT may include file contents. Files can be downloaded when Code Interpreter is enabled

- Disley_et_al_2015a.pdf**
PDF
- Disley_et_al_2011a.pdf**
PDF

Preview



The Peterborough SIB

This bot answers questions about the first social impact bond in the UK. Launched in 2010, the project aimed at reducing the reconviction rates of short-sentence male prisoners leaving HMP Peterborough.

Message The Peterborough SIB...

The importance of training data

We created a bot a provided it with evaluation reports, the INDIGO spreadsheets, case studies, and other sources of information from the GO Lab Library of Resources

Documents



[Peterborough social impact bond: background](#)

PDF, 246 KB, 3 pages



[Peterborough social impact bond: final report on cohort 2 analysis](#)

PDF, 117 KB, 18 pages



[Peterborough social impact bond: methodological review of the cohort 1 analysis](#)

PDF, 223 KB, 44 pages



[Peterborough social impact bond: learning exercise](#)

PDF, 111 KB, 17 pages



You

Can you please tell me how you have been trained?



The Peterborough SIB

I have been trained specifically to understand and provide information about the Peterborough Social Impact Bond (SIB) project, including details on its stakeholders, goals, outcomes, and any relevant data or findings from evaluations and reports related to the project. My training involves detailed analysis of documents, such as evaluation reports, related to the Peterborough SIB, which was an innovative financial mechanism aimed at reducing reoffending rates among short-sentenced offenders by funding interventions through private investments that would be repaid by the government if certain outcomes were achieved.

The key documents that form the basis of my training include the "Final Process Evaluation Report" of the Peterborough SIB pilot, conducted by RAND Europe, and provided by the Ministry of Justice. This report outlines the aims, operation, and outcomes of the SIB, including how it aimed to reduce reoffending through the One Service intervention, which provided support to offenders both in prison and after release. It also discusses the roles of various stakeholders involved in the project, including the Ministry of Justice, Social Finance, St Giles Trust, Sova, and others, as well as the innovative elements of the SIB model and its potential for replication in other contexts.

Training tailored GPT with HMP Peterborough



Benchmarking to information collected by human eye (on INDIGO Spreadsheet and website).

1ST TEST: Internet access and information accuracy:

Assumption 1: based on documents uploaded the GPT when prompted will be able to extract the same quality information from multiple documents as a human eye while reducing research time. - AMBIGUOUS

Assumption 1.1: even if the GPT has access to the web it will still rely mainly on the documents it's been provided with for data extraction. - YES

Assumption 1.2: it will need tailored prompts to better perform while reviewing the documents.- YES

Testing: With the same prompts the GPT did several rounds. Answers were compared to information on website and with each other.

Training tailored GPT with HMP Peterborough



Findings:

Turning on the internet did not have a great impact on the results.

- The quality of answers was influenced by the prompts of the rounds (eg.: focus on only documents while internet is turned on).
- **Satisfactory results for stage of development, policy sector, location related data.** Somewhat same, good results for SDG relation, service users/target population, commissioners/ outcome payers, intermediary organisations. Need improvement on outcome metrics, intervention description. **Difficult: capital raised** (could be due to not having a contract specifying this), **precise start date and contract signature** (months), **service providers and investors** (states that the documents do not have this information but when specifically asked if it can locate them using their name it finds them - when asked why it wasn't able to pick these up it asked for more detailed prompts).

Training tailored GPT with HMP Peterborough



2nd TEST: Prompt design with GPT and information accuracy:

Assumption 1: if the GPT is trained with the answers that a human eye gathered it will be able to write prompts for itself. - YES

Assumption 2: the GPT will be able to extract matching information with the prompts it generated for itself from the same project. - AMBIGUOUS

Testing:

A) Asking GPT to design 3 prompts knowing the answer (eg.: completed) and the type of information (eg.: stage of development) also not being allowed to use the answer words in the prompt.

B) Asking GPT to answer with the prompts it generated.

Training tailored GPT with HMP Peterborough



Findings:

It doesn't provide short answers (would need to be prompted to write in short).

- It performed well on some questions, while similarly or worse on others compared to first Test:
Effective and better than 1st Test on: contract signature, start of the project, outcome metrics
Same good performance on: policy sector, service users/target population, intervention description, location information, commissioners/ outcome payers, intermediary organisations
Didn't perform well: stage of development (worse than 1st test), capital raised (same as first test), service providers and investors (same as first test)
- Question about whether the project is completed: In its answer it went through its thought process and wasn't able to determine if it was completed. After this when asking it with prompt used in the 1st Test it was still unable to answer correctly.



GOVERNMENT OUTCOMES LAB

Feedback from Expert



@golaboxford



golab.ox.ac.uk

Next steps

Let's write a report ->



Next steps



Come to our social event! (same link, bring something INDIGO)

Thank you!

