How to find a novel idea?

How to find a new idea?

1. **Literature Review**: Conducted using Google Scholar to identify existing research and knowledge gaps.

2. **Idea Development**: Formulated based on the observation that no existing solutions effectively address the current challenges.

3. **Evaluation**: Systematic testing and validation of the proposed approach or method.

4. **Write-up**: Clear and structured documentation of findings, methodology, and implications.

1. **Literature Review**
   Example: LLM medical foundation model

2. **Idea Development**: Formulated based on the observation that no existing solutions effectively address the current challenges.

[HTML] Large language **models** in **medicine** [HTML] nature.com

AJ Thirunavukarasu, DSJ Ting, K Elangovan… - Nature **medicine**, 2023 - nature.com

… -art **LLM** applications in **medicine**, using ChatGPT as an illustrative example. First, **LLM** development is explained, outlining **model** … in developing these **models**. Next, the applications of …

☆ Save 99 Cite Cited by 2170 Related articles All 7 versions

The shaky **foundations** of large language **models** and **foundation models** for electronic health records [PDF] nature.com

M Wornow, Y Xu, R Thapa, B Patel, E Steinberg… - … digital **medicine**, 2023 - nature.com

… ChatGPT is a large language **model** (**LLM**), a type of FM which ingests text and outputs text in response. Though ChatGPT was trained to simply predict the next word in a sentence—it is …

☆ Save 99 Cite Cited by 169 Related articles All 10 versions

2023

Foundation models for generalist **medical** artificial intelligence [PDF] nature.com

M Moor, O Banerjee, ZSH Abad, HM Krumholz… - Nature, 2023 - nature.com

… GMAI, a class of advanced **medical foundation models**. 'Generalist' … **medical** applications, largely replacing task-specific **models**. Inspired directly by **foundation models** outside **medicine**…

☆ Save 99 Cite Cited by 1090 Related articles All 15 versions

A data-efficient strategy for building high-performing **medical foundation models**

Y Sun, W Tan, Z Gu, R He, S Chen, M Pang… - Nature Biomedical …, 2025 - nature.com

… **Medical foundation models** have recently emerged in several fields, … of **medical** artificial intelligence (AI) that can be applied to numerous downstream tasks. **Medical foundation models** …

☆ Save 99 Cite Related articles All 2 versions

A comprehensive survey of **foundation models** in **medicine** [PDF] ieee.org

W Khan, S Leem, KB See, JK Wong… - IEEE Reviews in …, 2025 - ieeexplore.ieee.org

Foundation models (FMs) are large-scale deeplearning **models** that are developed using large datasets and self-supervised learning methods. These **models** serve as a base for …

☆ Save 99 Cite Cited by 12 Related articles All 3 versions

Medical foundation large language **models** for comprehensive text analysis and beyond [PDF] nature.com

Q Xie, Q Chen, A Chen, C Peng, Y Hu, F Lin… - … Digital **Medicine**, 2025 - nature.com

… Table 1 compares the performance of our Me-LLaMA 13/70B **foundation models** against … Me-LLaMA 13B **model** surpassed the similar-sized **medical foundation model** PMC-LLaMA …

☆ Save 99 Cite Related articles All 4 versions

[HTML] Self-improving generative **foundation model** for synthetic **medical** image generation and clinical applications [HTML] nature.com

J Wang, K Wang, Y Yu, Y Lu, W Xiao, Z Sun, F Liu… - Nature **Medicine**, 2025 - nature.com

… In many clinical and research settings, the scarcity of high-quality **medical** imaging datasets … **medical** image–text generative **model** called MINIM that is capable of synthesizing **medical** …

☆ Save 99 Cite Cited by 9 Related articles All 2 versions

Fair **Foundation Models** for **Medical** Image Analysis: Challenges and Perspectives [PDF] arxiv.org

D Queiroz, A Carlos, A Anjos, L Berton - arXiv preprint arXiv:2502.16841, 2025 - arxiv.org

… **Foundation Models** (FMs), trained on vast datasets through self-supervised learning, enable … across **medical** imaging tasks while reducing dependency on labeled data. These **models** …

☆ Save 99 Cite Related articles All 2 versions »

Uncertainty of Vision **Medical Foundation Models** [PDF] openreview.net

H Huang, N Razavian - … and Hallucination in **Foundation Models**: The Next … - openreview.net

With recent emphasis on building domain specific visual **foundation model** in **medicine** (

2025 medical foundation model papers still can be accepted.

But the contribution become less important.

You need to try very hard to compete with other restaurants

Less competition

Less competition

1. If the contributions become small, it is still meaningful, but it just can not be accepted by tier 1 journals.

2. It needs a huge contribution to get accepted by tier 1 journal if you still working on an old topic. (You pizza is the No.1 delicious)

1. **Evaluation**: Systematic testing and validation of the proposed approach or method.

   Do more evaluations to prove that your idea is correct.

2. **Write-up**: Clear and structured documentation of findings, methodology, and implications.

   Write a clear, easy to read research papers.
   If you wrote a paper which is hard to follow, your paper also easy to be rejected.

Article

https://doi.org/10.1038/s41591-023-02504-3

A visual–language foundation model for pathology image analysis using medical Twitter

Received: 26 March 2023

Accepted: 18 July 2023

Published online: 17 August 2023

Check for updates

Zhi Huang[1,2,4], Federico Bianchi[3,4], Mert Yuksekgonul[3], Thomas J. Montine[2] & James Zou[1,3]

The lack of annotated publicly available medical images is a major barrier for computational research and education innovations. At the same time, many de-identified images and much knowledge are shared by clinicians on public forums such as medical Twitter. Here we harness these crowd platforms to curate OpenPath, a large dataset of 208,414 pathology images paired with natural language descriptions. We demonstrate the value of this resource by developing pathology language–image pretraining (PLIP), a multimodal artificial intelligence with both image and text understanding, which is trained on OpenPath. PLIP achieves state-of-the-art performances for classifying new pathology images across four external datasets: for zero-shot classification, PLIP achieves F1 scores of 0.565–0.832 compared to F1 scores of 0.030–0.481 for previous contrastive language–image pretrained model. Training a simple supervised classifier on top of PLIP embeddings also achieves 2.5% improvement in F1 scores compared to using other supervised model embeddings. Moreover, PLIP enables users to retrieve similar cases by either image or natural language search, greatly facilitating knowledge sharing. Our approach demonstrates that publicly shared medical information is a tremendous resource that can be harnessed to develop medical artificial intelligence for enhancing diagnosis, knowledge sharing and education.

The lack of annotated publicly available medical images is a major barrier for computational research and education innovations. At the same time, many de-identifed images and much knowledge are shared by clinicians on public forums such as medical Twitter (background, problem, and motivations). Here we harness these crowd platforms to curate OpenPath, a large dataset of 208,414 pathology images paired with natural language descriptions . We demonstrate the value of this resource by developing pathology language–image pretraining (PLIP), a multimodal artifcial intelligence with both image and text understanding, which is trained on OpenPath.(Solutions) PLIP achieves state-of-the-art performances for classifying new pathology images across four external datasets: for zero-shot classifcation, PLIP achieves F1 scores of 0.565–0.832 compared to F1 scores of 0.030–0.481 for previous contrastive language–image pretrained model. Training a simple supervised classifer on top of PLIP embeddings also achieves 2.5% improvement in F1 scores compared to using other supervised model embeddings. Moreover, PLIP enables users to retrieve similar cases by either image or natural language search, greatly facilitating knowledge sharing. (Results) **Our approach demonstrates that publicly shared medical information is a tremendous resource that can be harnessed to develop medical artifcial intelligence for enhancing diagnosis, knowledge sharing and education.** (Conclusions)

Example of our projects: Using blockchain for data sharing.

The solution is weak

Subscription-based **data-sharing** model **using blockchain** and **data** as a service [PDF] ieee.org
FA Al-Zahrani - Ieee Access, 2020 - ieeexplore.ieee.org
… The **blockchain**-based **data-sharing** models are gaining popularity as they make **data** …
A secure **data-sharing** model is proposed in [3] **using blockchain**. This model is specifically …
☆ Save 99 Cite Cited by 47 Related articles All 3 versions

Integrating **blockchain** for **data sharing** and collaboration in mobile healthcare applications [PDF] ieee.org
X Liang, J Zhao, S Shetty, J Liu… - 2017 IEEE 28th annual …, 2017 - ieeexplore.ieee.org
… systems, as well as the concept of self-sovereign **data** ownership, we propose an … **data sharing** solution by utilizing a decentralized and permissioned **blockchain** to protect privacy **using** …
☆ Save 99 Cite Cited by 849 Related articles All 4 versions

Blockchain as a notarization service for **data sharing** with personal **data** store [PDF] ieee.org
MJM Chowdhury, A Colman, MA Kabir… - … on big **data** science …, 2018 - ieeexplore.ieee.org
… Firstly, we present a architecture of **blockchain** based **data sharing** for Personal **Data** Storage (PDS), where **blockchain** … We developed a prototype **using** Ethereum private **blockchain**. …
☆ Save 99 Cite Cited by 114 Related articles All 4 versions

Blockchain based intelligent vehicle **data sharing** framework [PDF] arxiv.org
M Singh, S Kim - arXiv preprint arXiv:1708.0972…, 2017 - arxiv.org
… ubiquitous **data** access in a secure way. Our proposal is based on a very simple concept of **using Blockchain** based trust environment for **data sharing** among Intelligent Vehicles **using** …
☆ Save 99 Cite Cited by 187 Related articles All 4 versions »

Example of our projects: Remote monitoring and diagnosis

The solution is weak

A telemedicine system for **remote** cooperative **medical** imaging diagnosis
EJ Go, F Del Pozo, JA Quiles, MT Arredondo… - Computer Methods and …, 1996 - Elsevier
… [161, that assumes the **sharing** of the complete **remote** user screen, the WYSIRTWIS mode supports applications with several levels of **sharing**. This means that application screens can …
☆ Save 99 Cite Cited by 77 Related articles All 6 versions

Secure data **sharing** with blockchain for **remote** health monitoring applications: a review [PDF] springer.com
V Upadrista, S Nazir, H Tianfield - Journal of Reliable Intelligent …, 2023 - Springer
… Blockchain can harness the data stream to improve the quality of **remote** care provided by streamlining the **sharing** of **medical** records [23], protecting sensitive data from hackers, and …
☆ Save 99 Cite Cited by 34 Related articles All 8 versions

Conditional anonymous **remote** healthcare data **sharing** over blockchain [PDF] ieee.org
J Liu, W Jiang, R Sun, AK Bashir… - IEEE journal of …, 2022 - ieeexplore.ieee.org
… **remote medical** services and promote the development of the related healthcare industry. However, in traditional centralized data **sharing** … in the fully anonymous **sharing** schemes, we …
☆ Save 99 Cite Cited by 49 Related articles All 7 versions

Defining **remote medical** practice [PDF] wiley.com
JD Smith, SA Margolis, J Ayton, V Ross… - **Medical** Journal of …, 2008 - Wiley Online Library
… "**remote medicine**" or "**remote medical** practice" in Australia. A keyword database search (Box 2) identified 10 articles containing "**remote medical** … They usually **share** their workloads with …
☆ Save 99 Cite Cited by 59 Related articles All 14 versions

How to make your pizza become No 1?

1, Old solutions applied in a new problem and the result is prominent.

2, New solutions applied in an old probem.

3, New Solutions applied in a new problem (Breaking news, like ChatGPT)

Example:
Fraudulent prescriptions + Blokchain

Very interesting project
But not a novel idea. ✖

*******How to find the fraudulent prescriptions？

Fraudulent prescriptions

+

Use BERT model + Smart contracts

The result is the most accurate than any other solutions: 99% accuracy

Remote monitoring + message encrytion, etc

Very hot topic

But not novel: old questions with old solutions ✖

Remote monitoring + Small LLMs models to automatically detecting patient conditions + malicious behaviors

Small device with AI medical model detection

CSC 116 Data Anonymization

Using a small local LLM (llama 3.1 8B) to anonymize data for a remote big LLM (ChatGPT/Claude Sonnet)

Sacha Storz · Follow
5 min read · Aug 11, 2024

80

If you have data that includes sensitive information like names or other personal details, it's probably best not to send it to a remote LLM like ChatGPT or Claude Sonnet. With GDPR and other privacy regulations in play, it's better to be safe than sorry.

For instance, I often use these powerful LLMs to get well-structured summaries of interviews I conduct with clients. In the past, I would manually anonymize the interviews, replacing all names (whether of people or companies) with placeholders. Then, I would upload the anonymized content to ChatGPT, along with instructions on how to work with the material.

Learn Presidio
Home
Concepts
Tutorial
Text de-identification
Home
Presidio Analyzer
Presidio Anonymizer
Home
Developing PII anonymization operators
Image de-identification
Structured and Semi-structured
PII detection evaluation

Presidio Anonymizer

The Presidio anonymizer is a Python based module for anonymizing detected PII text entities with desired values. Presidio anonymizer supports both anonymization and deanonymization by applying different operators. Operators are built-in text manipulation classes which can be easily extended.

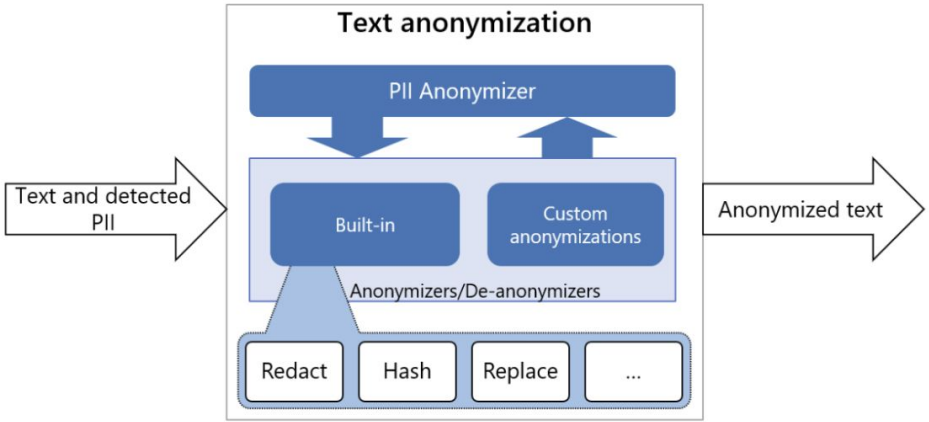

Table of contents
Installation
Getting started
Using docker container
Using python runtime
Main concepts
Built-in operators
Handling overlaps between entities
Additional examples for overlapping PII scenarios
Creating a new operator
API reference

Text Anonymization

TEXT ANONYMIZATION | IMAGE ANONYMIZATION

Input text

Here are a few examples sentences we currently support:

Hello, my name is David Johnson and I live in Maine.
My credit card number is 4095-2609-9393-4932 and my crypto wallet id is 16Yeky6GMjeNkAiNcBY7ZhrLoMSgg1BoyZ.

On September 18 I visited microsoft.com and sent an email to test@presidio.site, from the IP 192.168.0.1.

My passport: 191280342 and my phone number: (212) 555-1234.

This is a valid International Bank Account Number: IL150120690000003111111 . Can you please check the status on bank account 954567876544?

Kate's social security number is 078-05-1126. Her driver license? it is 1234567A.

Anonymized text

Here are a few examples sentences we currently support:

Hello, my name is <PERSON> and I live in <LOCATION>.
My credit card number is <CREDIT_CARD> and my crypto wallet id is <CRYPTO>.

On <DATE_TIME> I visited <URL> and sent an email to <EMAIL_ADDRESS>, from the IP <IP_ADDRESS>.

My passport: <US_PASSPORT> and my phone number: <PHONE_NUMBER>.

This is a valid International Bank Account Number: <IBAN_CODE> . Can you please check the status on bank account <US_BANK_NUMBER>?

<PERSON>'s social security number is <US_SSN>. Her driver license? it is <US_DRIVER_LICENSE>.

Thanks!