

Scalable Patent Search and Analysis Using LLMs and Function Calling

RMACC Symposium 2026



Juanjo García Mesa
Research Software Engineer
Arizona State University



Outline



Motivation

The Data

Retrieval-Augmented Generation

Building the Database

Query Workflow

Demo

Collaborators



Jordan Bell



Gil Speyer



Johnathan Lee



Roger Kohler



Faraj Tessilimi

Motivation



SI manages 4,000 ASU granted patents and applications



SI manages 4,000 ASU granted patents and applications



Manual classification and queries are slow

I have an investor who invests in the sustainability space. The investor wants to see all ASU technologies that fall into the sustainability space. Please produce a list that includes patents and patent applications by number, title, and includes a one-paragraph summary

I have a
sustain
to see a
the sus
a list th
applicat
include

List all the granted patents and patent applications by inventor Erica Forzani

I have an
sustained
to see a
the sus
a list th
applicat
include

List all t
applicat

Summarize the composition of ASU's
issued United States patents by
technical subject matter in plain English



SI manages 4,000 ASU granted patents and applications



Manual classification and queries are slow



Request: Accurate, fast, natural-language-driven tool



SI manages 4,000 ASU granted patents and applications



Manual classification and queries are slow



Request: Accurate, fast, natural-language-driven tool



“Of course, it’s definitely possible”

The Data



</Open Data Portal>

Home

Data products ▾

APIs ▾

Support ▾

MyODP

Register for ODP access

OPEN DATA PORTAL
API

Getting started

API syntax examples

API rate limits

Patent File Wrapper
API ▾

Bulk Datasets API ▾

Final Petition
Decisions API ▾

PTAB Trials API ▾

PTAB Appeals API ▾

PTAB Interferences
API ▾

Office Action Text
Retrieval ▾

[Home](#) > [APIs](#) > Getting started

Getting started

What is the ODP API?

The Open Data Portal (ODP) API allows you to access USPTO data at no cost – with several ways to do it.

The ODP API uses HTTP, GET, and POST requests to return responses in JSON format. ODP also provides REST APIs, which provide rich search capabilities and allow users to retrieve patent application data, continuity data, documents, patent term adjustment data, and much more. You must get an API Key to use the ODP API.

Note: The data products on our website allow users to download data without the need for an API key. Using ODP APIs will require an API key.

Resources for APIs:

- The [API Syntax Examples](#) section is helpful for a “quick start” and may be used by more casual users to flexibly search the same data available on the Patent File Wrapper.

The [Support UI E7](#) has examples, schemas, and more to test out ODP APIs for Patent File Wrapper and Bulk

ASU's Patent Assignees

"Arizona Board of Regents on Behalf of Arizona State University"

"ARIZONA BOARD OF REGENTS ON BEHALF OF ARIZONA STATE UNIVERSITY"

"Arizona Board of Regents on behalf of Arizona State University"

"[The] Arizona Board of Regents"

"**Arizone** State", "Ariz. State", "ASU", "**ArIzona** State University", "Arizona Board of **Regenta**"

Patent Application Full-Text Data (No Images)

Product Identifier: APPXML

Provides the bulk zip files that contain the concatenated full-text of each patent application XML document (non-provisional utility and plant). This page provides an additional feature called "View Patent Records" which allows user to find or discover the patent applications that are bundled in the zip file. However, in this zip file, there may be applications that are not available in ODP and will not be included in the search results. Also, if there are revisions to the applications, the file will only show the original application.

Dates available: March 15, 2001 - July 17, 2025 | **Last Updated:**

Start Date:

End Date:

01-01-2002



07-19-2025



Filter

Reset

API Query

January 1, 2002 - July 19, 2025

Showing 1 to 25 of 1266 Results

File Name	Release Date
ipa250717.zip	2025-07-17 / 3:59:45 AM EST
ipa250710.zip	2025-07-10 / 3:59:43 AM EST
ipa250703.zip	2025-07-03 / 4:00:14 AM EST
ipa250626.zip	2025-06-26 / 4:00:03 AM EST
ipa250619.zip	2025-06-19 / 3:01:58 AM EST

Patent Grant Full-Text Data (No Images) - XML

Product Identifier: PTGRXML

Provides the bulk zip files that contains the concatenated full-text of each patent grant document issued weekly. This page provides an additional feature called "View Patent Records" which allows user to find or discover the patent grants that are bundled in the zip file. Even though the zip file may contain grants that belong to the applications which were filed before 2001, ODP will only show the grants that belong to the applications that were filed from 2001.

Dates available: January 1, 2002 - July 15, 2025 | **Last Updated:** 2025-07-15 (Weekly - Tuesday)

Start Date:

End Date:

01-01-2002



07-19-2025



Filter

Reset

API Query

January 1, 2002 - July 19, 2025

Showing 1 to 25 of 1238 Results

File Name	Release Date	File Size (Bytes)	Actions
ipg250715.zip	2025-07-15 / 3:57:53 AM EST	171,350,746	View Patent Records
ipg250708.zip	2025-07-08 / 3:57:53 AM EST	169,543,976	View Patent Records
ipg250701.zip	2025-07-01 / 3:57:52 AM EST	158,272,951	View Patent Records
ipg250624.zip	2025-06-24 / 3:01:41 AM EST	138,834,644	View Patent Records
ipg250617.zip	2025-06-17 / 3:01:42 AM EST	173,536,353	View Patent Records

Documents and Resources

[us-patent-grant-v47-2022-02-17.dtd](#)
[us-sequence-listing-v1_3-2020-10-08.dtd](#)
[us-patent-grant-v46-2021-08-30.dtd](#)
[us-patent-grant-v45-2014-04-03.dtd](#)
[us-patent-grant-v44-2013-05-16.dtd](#)
[us-patent-grant-v43-2012-12-04.dtd](#)
[us-patent-grant-v42-2006-08-23.dtd](#)
[us-patent-grant-v40-2004-12-02.dtd](#)
[us-patent-grant-v41-2005-08-25.dtd](#)

```
Vaults SFTP X Sol login +
1 <?xml version="1.0" encoding="UTF-8"?>
2 <!DOCTYPE PATDOC SYSTEM "ST32-US-Grant-025xml.dtd" [
3 <!ENTITY USD0452599-20020101-D00000.TIF SYSTEM "USD0452599-20020101-D00000.TIF" NDATA TIF>
4 <!ENTITY USD0452599-20020101-D00001.TIF SYSTEM "USD0452599-20020101-D00001.TIF" NDATA TIF>
5 <!ENTITY USD0452599-20020101-D00002.TIF SYSTEM "USD0452599-20020101-D00002.TIF" NDATA TIF>
6 ]>
7 <PATDOC DTD="2.5" STATUS="Build 20020101">
8 <SDOB1>
9 <B100>
10 <B110><DNUM><PDAT>D0452599</PDAT></DNUM></B110>
11 <B130><PDAT>S1</PDAT></B130>
12 <B140><DATE><PDAT>20020101</PDAT></DATE></B140>
13 <B190><PDAT>US</PDAT></B190>
14 </B100>
15 <B200>
16 <B210><DNUM><PDAT>29129833</PDAT></DNUM></B210>
17 <B211US><PDAT>29</PDAT></B211US>
18 <B220><DATE><PDAT>20000922</PDAT></DATE></B220>
19 </B200>
20 <B300>
21 <B310><DNUM><PDAT>400 02 946</PDAT></DNUM></B310>
22 <B320><DATE><PDAT>20000322</PDAT></DATE></B320>
23 <B330><CTRY><PDAT>DE</PDAT></CTRY></B330>
24 </B300>
25 <B400>
26 <B472>
27 <B474><PDAT>14</PDAT></B474>
28 </B472>
29 </B400>
bulk_ptgrxml_unzip/pg020101.xml
```

```
Vaults SFTP X Sol login +
1 <?xml version="1.0" encoding="UTF-8"?>
2 <!DOCTYPE us-patent-grant SYSTEM "us-patent-grant-v40-2004-12-02.dtd" [ ]>
3 <us-patent-grant lang="EN" dtd-version="v40 2004-12-02" file="USD0500396-20050104.XML" status="SAMPLE-D
3 ATA" id="us-patent-grant" country="US" date-produced="20041221" date-publ="20050104">
4 <us-bibliographic-data-grant>
5 <publication-reference>
6 <document-id>
7 <country>US</country>
8 <doc-number>D0500396</doc-number>
9 <kind>S1</kind>
10 <date>20050104</date>
11 </document-id>
12 </publication-reference>
13 <application-reference appl-type="design">
14 <document-id>
15 <country>US</country>
16 <doc-number>29198423</doc-number>
17 <date>20040129</date>
18 </document-id>
19 </application-reference>
20 <us-application-series-code>29</us-application-series-code>
21 <priority-claims>
22 <priority-claim sequence="01" kind="national">
23 <country>JP</country>
24 <doc-number>2003-028833</doc-number>
25 <date>20031001</date>
26 </priority-claim>
27 </priority-claims>
28 <us-term-of-grant>
bulk_ptgrxml_unzip/ipg050104.xml
```

Data Parsing and Formatting

```
1 {
2   "format": "us-patent-application",
3   "application_info": {
4     "application_number": "15125024",
5     "country": "US",
6     "filing_date": "2015-03-10",
7     "series_code": "15",
8     "application_type": "utility"
9   },
10  "publication_info": {
11    "publication_number": "20170021009",
12    "kind_code": "A1",
13    "publication_date": "2017-01-26",
14    "country": "US"
15  },
16  "title": "Heat Inactivated Poxvirus Improves Vaccination Resu
17  "abstract": "Compositions including a heat-inactivated poxvir
18  "description": "This application claims priority to U.S. Prov
19  "inventors": [
20    {
21      "name": "Bertram Jacobs",
22      "type": "individual",
23      "first_name": "Bertram",
24      "middle_name": "",
US-20170021009-A1.json

1 {
2   "patent_info": {
3     "patent_number": "11978466",
4     "kind_code": "B2",
5     "publication_date": "2024-05-07",
6     "country": "US",
7     "application_number": "17827438",
8     "application_date": "2022-05-27",
9     "date_produced": "2024-04-16"
10  },
11  "title": "Systems, methods, and apparatuses for restoring degraded speech via a modified diffusion model",
12  "abstract": "Systems, methods, and apparatuses to restore degraded speech via a modified diffusion model are descri
13  "description": "This application is related to, and claims priority to, U.S. Provisional Patent Application No. 63/
14  "inventors": [
15    {
16      "name": "Jianwei Zhang",
17      "first_name": "Jianwei",
US-11978466-B2.json

2,1 Top
```



Retrieval-Augmented Generation

Leverage CPC codes

G	06	C	7	02
Section	Class	Subclass	Main group	Subgroup

G Physics

06 Computing; calculating or counting

C Digital computers in which all the computation is effected mechanically

7 Input mechanisms

02 Keyboards

Leverage CPC codes

G	06	C	7	02
Section	Class	Subclass	Main group	Subgroup

FAIL

G Physics

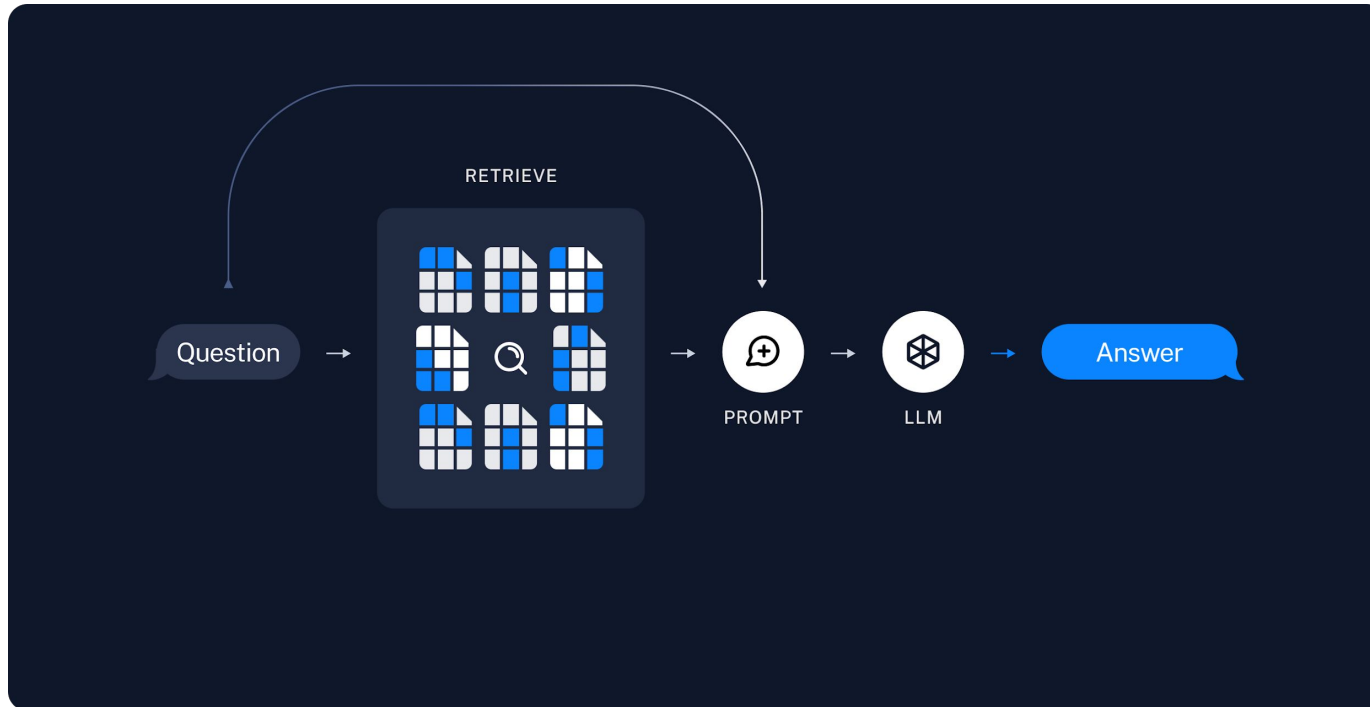
06 Computing, calculating, or counting

C Digital computers in which all the computation is effected mechanically

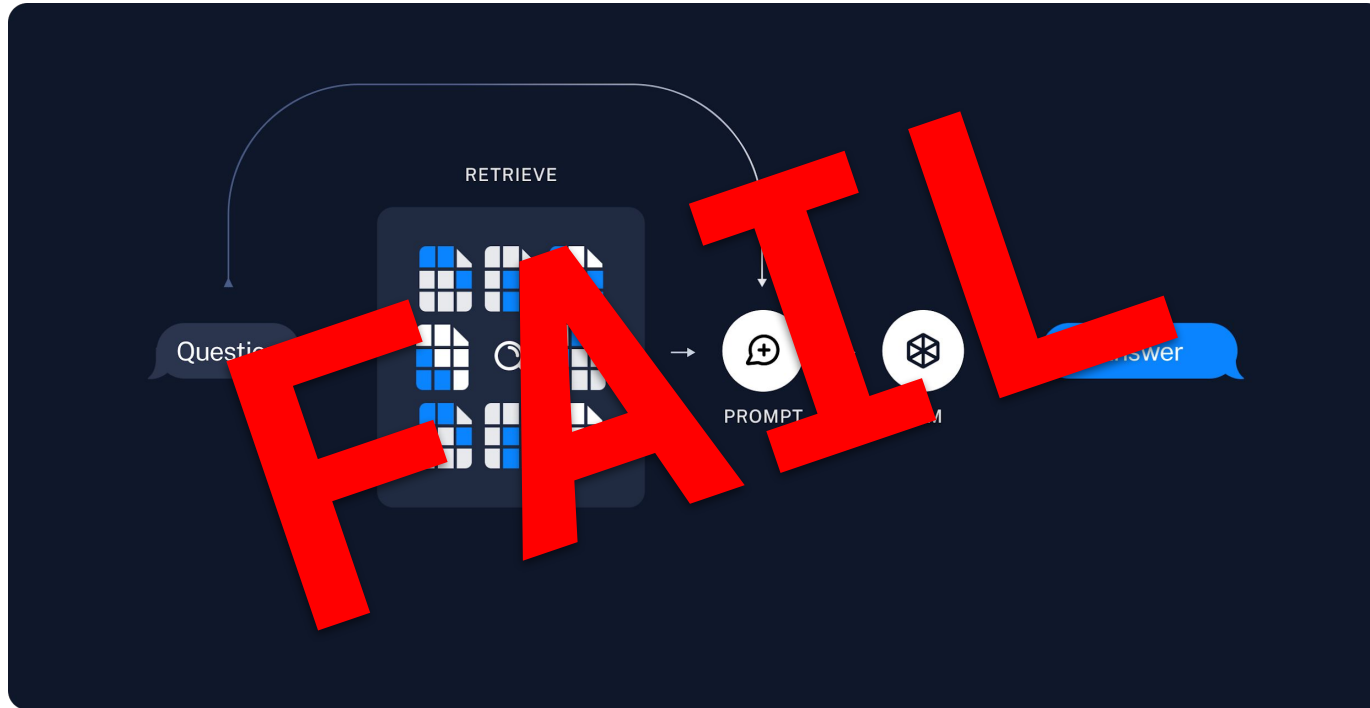
7 Input mechanisms

02 Keyboards

Traditional Retrieval-Augmented Generation



Traditional Retrieval-Augmented Generation





SOME TIME LATER...

Multi-stage Pipeline



Build the Database

PostgreSQL + pgvector Embeddings

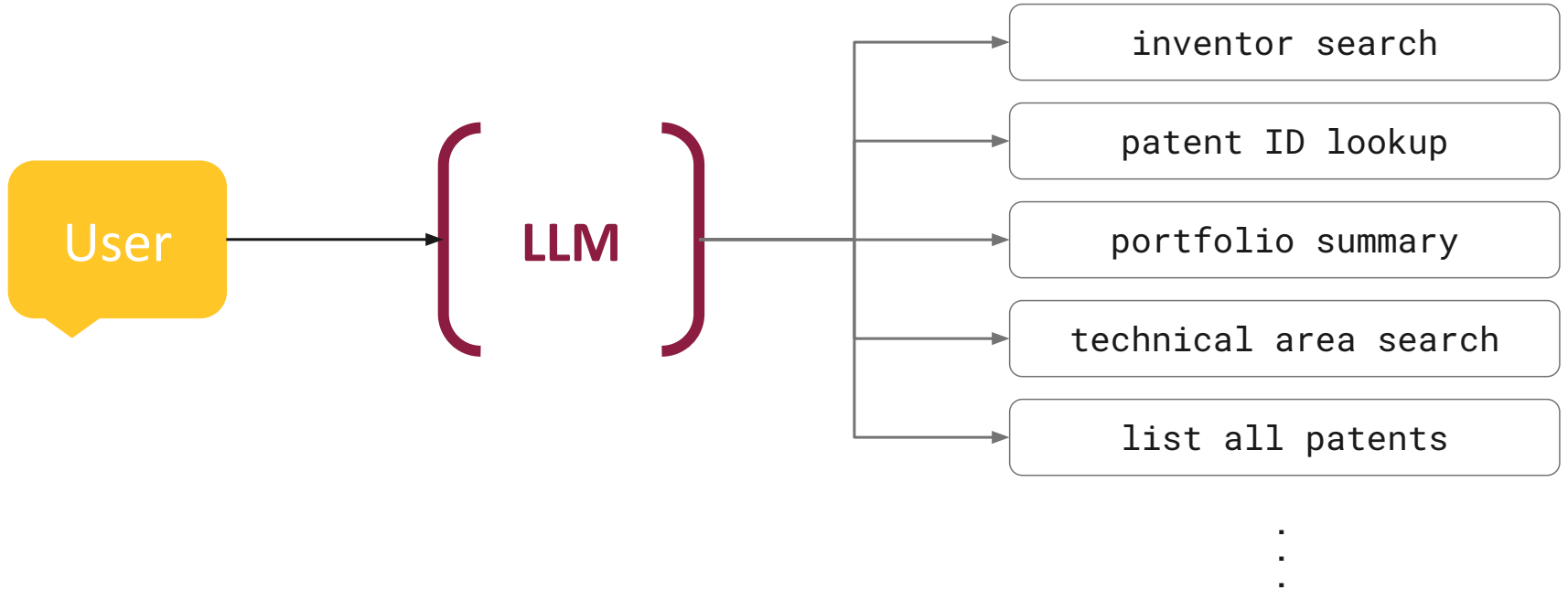
Each JSON file is added to the database

- title, abstract, claims, inventors, assignees
- classifications: a dict of CPC code
- searchable text: title + abstract + claims text + CPC codes
- technical areas: list of 1-3 LLM pre-generated taxonomy slugs
- embedding: pgvector embedding using searchable text
<https://github.com/pgvector/pgvector/>



Query Workflow

Function Calling



```
Vaults SFTP Local Terminal
1 {
2   "type": "function",
3   "function": {
4     "name": "technical_area_search",
5     "description": "Find patents that fall under a specific category.",
6     "parameters": {
7       "type": "object",
8       "properties": {
9         "categories": {
10          "type": "array",
11          "items": {"type": "string"},
12          "description": "List of categories to search for patents."
13        },
14      },
15      "required": ["categories"],
16    },
17  },
18 }
19
```

[No Name] [+]

The Berkeley Function Calling Leaderboard (BFCL): From Tool Use to Agentic Evaluation of Large Language Models

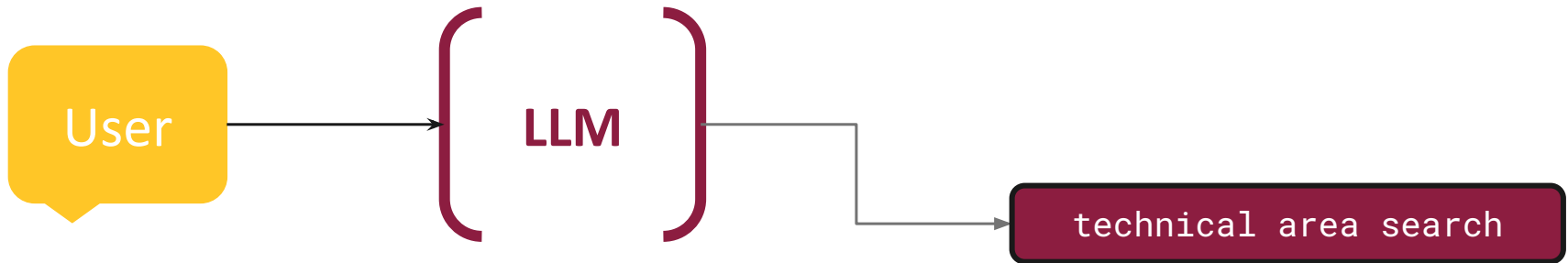
Shishir G. Patil¹ Huanzhi Mao¹ Fanjia Yan¹ Charlie Cheng-Jie Ji¹ Vishnu Suresh¹
Ion Stoica¹ Joseph E. Gonzalez¹

Abstract

Function calling, also called tool use, refers to an LLM's ability to invoke external functions, APIs, or user-defined tools—an essential capability for agentic LLM applications. Despite its prominence, there does not exist a standard

soning, and creative multimodal tasks. However, agents powering use-cases such as coding, and knowledge discovery (Huang et al., 2024; Yao et al., 2022) often require LLMs to interface with external (e.g., web-search) to either retrieve up-to-date information or to enact actions which have real-world consequences. To aid in this, previous work such as GPT4o (OpenAI, 2024) and Toolformer (Schick et al., 2023)

Patents Related to a Technical Area



Hybrid Retrieval Score (more art than science)

```
score = lexical * 0.4 + semantic * 0.4 + taxonomy * 0.15 + cpc * 0.05
```

- lexical: overlap between augmented query keywords and searchable text
- semantic: cosine similarity between query embedding and record embedding
- taxonomy: predicted query taxonomy overlap with technical areas and taxonomy keywords that appear in searchable text
- cpc: record CPC codes matching CPC prefixes from predicted taxonomy areas

```
1 version: "2026-04-09"
2 areas:
3   - slug: therapeutics-drug-delivery
4     name: Therapeutics & Drug Delivery
5     description: Drug candidates, formulations, therapeutic compositions, and delivery
6     keywords: ["therapeutic", "drug delivery", "pharmaceutical", "treatment", "formula
7     cpc_prefixes: ["A61K", "A61P", "C07K", "C07D"]
8     example_themes: ["cancer therapy", "drug carrier", "peptide therapeutic"]
9   - slug: diagnostics-biosensing
10    name: Diagnostics & Biosensing
11    description: Detection, assays, biomarkers, biosensors, and analytical diagnostic
12    keywords: ["diagnostic", "biosensor", "assay", "detection", "biomarker", "analyte"
13    cpc_prefixes: ["G01N", "C12Q", "A61B"]
14    example_themes: ["point-of-care test", "multiplex assay", "electrochemical biosens
15  - slug: medical-devices-health-platforms
16    name: Medical Devices & Health Platforms
17    description: Medical devices, digital health systems, wearables, and procedural to
18    ols.
19    @
data/config/taxonomy.yaml 3,1 Top
```

Deployment

- Python FastAPI backend serving query/report endpoints
- Streamlit frontend for the chat
- Dockerized PostgreSQL database using the pgvector extension

- Currently exploring permanent deployment solutions

Demo

Sessions

New Session

New session Updated 2026-05-11 | 0 messages

Delete Session

API status: ready

ASU | SKYSONG INNOVATIONS

Patent Assistant

Ask, review evidence, inspect records, and export the run.

Start with a patent number, inventor, topic, or portfolio question.

Good first queries:

Summarize the composition of ASU's patent portfolio

List all patents by Erica Forzani

Find patents related to biosensors

Enter a patent number or ask a question about ASU's patent portfolio

I



Sessions

New Session

New session Updated 2026-05-11 | 0 messages

Summarize the composition of ASU's patent por... Updated 2026-05-11 | 2 messages

Delete Session

API status: ready

ASU | SKYSONG INNOVATIONS

Patent Assistant

Ask, review evidence, inspect records, and export the run.

Start with a patent number, inventor, topic, or portfolio question.

Good first queries:

Summarize the composition of ASU's patent portfolio

List all patents by Erica Forzani

Find patents related to biosensors

Enter a patent number or ask a question about ASU's patent portfolio

I



Sessions

New Session

New session Updated 2026-05-11 | 0 messages

List all the patents by Erica Forzani. Updated 2026-05-11 | 2 messages

Summarize the composition of ASU's patent por... Updated 2026-05-11 | 2 messages

Delete Session

API status: ready

ASU | SKYSONG INNOVATIONS

Patent Assistant

Ask, review evidence, inspect records, and export the run.

Start with a patent number, inventor, topic, or portfolio question.

Good first queries:

Summarize the composition of ASU's patent portfolio

List all patents by Erica Forzani

Find patents related to biosensors

Enter a patent number or ask a question about ASU's patent portfolio

Future Work

- Internal deployment to SI for testing
- Public access for university partners
- Polish report generation
- Add general conversation skills
- Identify missing functionalities
- Decouple the tool from the data and create a blueprint for other applications

Acknowledgements



Jordan Bell



Gil Speyer



Johnathan Lee



Roger Kohler



Faraj Tessilimi

