



www.ngisearch.eu

BOOKLET 2025

50 projects funded
between 2022 and 2025

Rev 1.0



This project has received funding from the European Union's Horizon Europe research and innovation programme under the grant agreement 101069364 and it is framed under Next Generation Internet Initiative.

AALLM Adversarial audit of LLM-based search engines	3
ADITIV (Asset Discovery In The Internet of Value)	4
AI-Generated Code Search	5
ALLMA	6
Better Food Search	7
carbon txt	8
Chat-EUR-Lex	9
ChatDKG	10
COGNITREK	11
COPS	12
Data space search engine EDC	13
Datami PRO	14
Debunker-Assistant	15
E V A Gallery - European Visual Arts Gallery	16
EGGO	17
Ensuring Fairness in Democratic AI	18
Fediverse Discovery Providers	19
FörderFunke	20
HeReFaNMI	21
InDi	22
IoT twinning for digital product passports	23
LabDiscoveryEngine	24
Ma Dada	25
Ma Dada UX	26
MindBugs Discovery	27
MOUSSE (Metadata fOcUsed Semantic Search Engine)	28
Neural Datafari	29
OKLLM: Online Knowledge Search for LLM innovations	30
On My Disk: search integration	31
On My Disk: Web UI	32
Open Data Deep Search	33
Open Politics	34
OSCAR4US	35
OSE	36
Qlever	37
Science Checker - reloaded	38
SCION Browser	39
Search-a-licious	40
Spare Scores	41
SVP64	42
SWH Scanner	43
TalentLayer	44
TETYS	45
The World Literature KG	46
Trallie	47
Trust4AI	48
TrustSearch	49
Verif ai	50
Visio Creativia	51
WAISE	53
WASPer	53

AALLM (Adversarial audit of LLM-based search engines)

LLM-based search engines produce content with factual errors and need to be regulated

Recent performance leap in Large Language Models (LLMs) led to build chatbots and search engines relying on them. These tools are widely used in this society. However, people are not always aware that the answers to critical questions they ask can be erroneous or biased. AI Forensics aims at providing proof of this undesirable behaviors to hold big tech companies releasing these tools accountable. Our focus will be to analyze Microsoft Copilot's answers on european elections' questions.



Repository:

<https://github.com/aiforensics/alexlab-intervention-factory-ce>

Team:

Sonia Tabti, Marc Faddoul, Salvatore Romano, Riccardo Angius

Country: France, Italy, Spain



Contact Email:

info@aiforensics.org

Other Info:

More Information: The github link will be available later.

ADITIV

Asset Discovery In The Internet of Value

Blockchain is increasingly gaining importance not only as a siloed technology for implementing on-chain dApps but also as an infrastructure element in the global internet-based computing landscape. In particular, the ability to uniquely identify assets allows easier implementation of traceability and auditability in any tokenisation process within the Internet of Value, thus increasing trust, privacy, and transparency regarding products and processes involving physical or digital assets. This grant proposal focuses on the development of an open “search engine indexing mechanism” that leverages digital identifiers and asset metadata whose states are bilaterally synchronised between on-chain and off-chain systems. ADITIV provides an open-source “search engine indexing mechanism” that leverages digital identifiers and asset metadata whose states are bilaterally synchronised between on-chain and off-chain systems. Uni Systems is now a strategic investor supporting the growth of Compellio, a software company based in Luxembourg with Denis Avrilion as CEO.

**Repository:**

<https://compellio.com>

Team:

compellio

Country: Luxembourg

**Contact Email:**

info@compellio.com

AI-Generated Code Search

Search, detect, and identify AI-generated code

AI-Generated Code Search enhances trust for users searching for code on the internet, especially for both AI-generated code found on the web and code created by a generative AI engine or LLM. This tool provides information and validation on the derivative open source origin of AI-generated code to identify its vulnerabilities and licenses to mitigate any software supply chain integrity or security risks associated with using AI-generated code.



Repository:

<https://github.com/aboutcode-org/ai-gen-code-search>

Team:

AboutCode

Country: Belgium



Contact Email:

pombredanne@gmail.com

Other Info:

<https://aboutcode.org/>

ALLMA: an AI-based LLM for Android

An /e/OS LLM Assistant experiment

The ALLMA project is:\n- a comprehensive comparison of Open Source LLM models\n- an experiment to run an Open Source LLM locally on a smartphone (offline) running /e/OS



Repository:

<https://gitlab.e.foundation/e/os/open-source-llm-assistant>

Team:

e Foundation

Country: France



Contact Email:

support@e.email

Other Info:

The alpha test concluded that while running an open-source LLM locally on a smartphone is possible, current hardware limitations create latency that significantly impacts user experience. With anticipated advancements in smartphone technology, local AI processing will likely become more viable, but in the meantime, a server-based solution is recommended.

Better Food Search

Search food products by properties you care about

Better Food Search (opengoods.org) allows users to search for specific ingredients, such as peanut butter, leveraging structured data and categories to provide detailed results.



Repository:

<https://gitlab.com/peerdb/peerdb>

Team:

Layer8 institute

Country: Slovenia



Contact Email:

info@plast8.si

Other Info:

<https://opengoods.org>

Carbon.txt

Sustainability information that is easier to discover and use

Carbon.txt proposes a single place to look on any domain - /carbon.txt - for public machine-readable sustainability data relating to that company. This standardised, distributed approach uses existing internet governance structures and industry norms to efficiently surface data that companies have to publish according to national laws. We envisage a data ecosystem growing around this that will support action on the defining crisis of our time - the climate crisis.



Repository:

<https://github.com/thegreenwebfoundation/carbon.txt>

Team:

Green Web Foundation - Chris Adams, Hannah Smith and Fershad Irani

Country: Germany, Netherlands



Contact Email:

support@thegreenwebfoundation.org

Other Info:

Our initial focus is on data relating to green energy claims.

<https://carbontxt.org/>

<https://www.thegreenwebfoundation.org/tools/carbon-txt/>

Chat-EUR-Lex

Trustworthy Generative AI

Chat-EUR-Lex develops a trustworthy chatbot interface to access EU legal documents, providing simplified explanations, and fostering users' understanding of EU law.



Repository:

<https://github.com/Aptus-AI/chat-eur-lex/>

Team:

Aptus.AI: Andrea Bolioli, Lorenzo De Mattei;
IGSG-CNR: Manola Cherubini, Francesco Romano

Country: Italy



Contact Email:

chat-eur-lex@igsg.cnr.it

Other Info:

<https://www.aptus.ai/>

<https://www.igsg.cnr.it/en/>

ChatDKG

An open framework for trusted AI based on OriginTrail DKG

ChatDKG is an open framework for efficient search, access, and utilization of trusted knowledge with AI, based on OriginTrail Decentralized Knowledge Graph (DKG), Paranets and Knowledge Assets. The ChatDKG framework is designed as an open community project to explore a diverse range of possible applications.



Repository:

<https://github.com/OriginTrail/ChatDKG>

Team:

Trace Labs

Country: Slovenia, Serbia



Contact Email:

branimir.rakic@origin-trail.com

Other Info:

<https://origintrail.io/>

<https://tracelabs.io/>

COGNITREK (Driving Cognitive Accessibility)

Driving Cognitive Accessibility

Cognitrek is an AI-powered tool that simplifies complex documents and adapts them to students' needs. It extracts and transforms information to make knowledge more accessible, including features that support individuals with learning difficulties. The platform ensures transparency, traceability, and complies with EU ethical and legal standards for AI, including data protection regulations.



Repository:

<https://github.com/XIMDEX/cognitrek>

Team:

Ximdex develops AI-driven platforms for smart content management and personalized education.

Country: Spain



Contact Email:

info@ximdex.com

Other Info:

<https://www.cognitrek.ximdex.net>

COPS

Confidential Private Search

COPS addresses the problem of private search by combining the functionality of the OpenSearch (<https://opensearch.org/>) search and analytics framework and the security primitives of confidential computing. The innovation behind this will enable data owners to maintain verifiable control over their data and ensure the protection of data in use.

**Repository:**

<https://github.com/canarybit/COPS>

Team:

Nicolae Paladi, Stefano Cucchiella

Country: Sweden**Contact Email:**

hi@canarybit.eu

Other Info:

<https://canarybit.eu>

Data space search engine (EDC)

Extension of the EDC project to add discoverability of data in the European data spaces

Today, companies are only able to share data with each other on a one-to-one basis. To share data, they have to copy-paste it using the Eclipse Data Components (EDC) Connector.

We're extending the EDC Connector to enable indexing, search and discovery of decentralized data among B2B partners. This will bring down the cost of sharing data to a couple of minutes per dataset, will enhance sovereignty over the shared data and will enable B2B data discovery.



Repository:

<https://git.startinblox.com/applications/ngi-search/indexing-service>

Team:

Startin'blox

Country: France



Contact Email:

sylvain@startinblox.com

Other Info:

<https://github.com/eclipse-edc/Connector/>

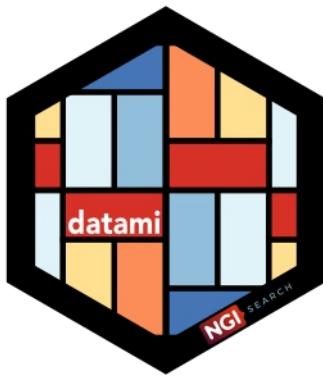
<https://git.startinblox.com/framework/sib-core/>

<https://startinblox.com/>

Datami

Making open data affordable for everyone

Datami PRO empowers small organizations to publish datasets as shareable visualizations that communities can contribute to. Built on the free Datami project, it will be a no-code, open-source platform allowing users to structure datasets into data packages, customize web components (maps, charts, cards, filters), manage contributions, and publish to open data catalogs. Datami PRO's affordability, flexibility, and respect of standards will make it the simplest tool for anyone's open data project.



Repository:

<https://gitlab.com/multi-coop/datami-project>

Team:

tech cooperative multi.coop, Julien PARIS
(lead dev)

Country: France



Contact Email:

datami@multi.coop

Other Info:

Datami website : <https://datami.multi.coop/?locale=en>

Debunker-Assistant

D-A is a citizen-based AI tool which supports the analysis and detection of online misinformation

The tool takes the link of a news article and returns its misinformation profile based on NLP and NA features, part of which are designed by involving non-expert citizens. D-A works with Italian and English news, and it is not bounded on a specific topic: it is designed as a general purpose tool that can extract relevant features for assessing the quality of information.



Repository:

<https://github.com/aequa-tech/debunker-assistant>

Team:

Marco Antonio Stranisci, Mirko Lai,
Giovanni Messina, Simona Frenda

Country: Italy



Contact Email:

simona.frenda@aequa-tech.com

Other Info:

Our project addresses the problem of misinformation spread on the internet. The solution is research-based (the team includes academic researchers) and focused on end-users. The software will be open source.

Two main topics:

- Ethics in search and discovery: specific tools and indicators to deal with misinformation, de-biasing mechanisms, explainable AI, trustworthy future internet.
- Network analysis: to gain insights about a certain context and aid the search and discovery of information.

E.V.A.Gallery - European Visual Arts Gallery

Europe's first AI-Powered Web3 Gallery for professional Artists and Gallerists.

European Visual Arts Gallery (E.V.A. Gallery) is a European wide research project, exploring latest art and cutting edge online tech. Artists and gallerists can make virtual exhibitions or create entire virtual galleries to present art in 3D online space. They can use AI tools to protect their art from plagiarism, also mint, exhibit and sell their own NFTs. The project is aimed at new ways of presentation and protection of visual data within the next generation internet.



Repository:

<https://github.com/eva-gallery>

Team:

EUforion, FIIT STU

Country: Slovakia



Contact Email:

info@evagallery.eu

Other Info:

<https://evagallery.eu/>

Enhanced Geocoder for gvSIG Online: Improving Geospatial Address Search

Open source modular geocoding service for interoperable and customizable spatial infrastructures

EGGO is an open source geocoding microservice designed for integration with gvSIG Online, enabling organizations to set up their own customizable and interoperable geocoding infrastructure. EGGO supports multiple geocoding engines (e.g., Nominatim, Photon, Google) and complies with open standards for spatial data infrastructures, empowering users with enhanced control, performance, and data sovereignty.



Repository:

<https://github.com/gvSIGAssociation/gvsig-online>

Team:

SCOLAB

Country: Spain



Contact Email:

info@gvsig.com

Other Info:

<https://www.gvsig.com/en>

<https://blog.gvsig.org/2023/06/26/geocoder-of-gvsig-online-versatility-and-power-in-address-searches/>

Ensuring Fairness in Democratic AI

Evaluating the impacts of public-focused decision-making systems based on actual benchmark data.

Matt Stempeck and Eticas are working together to help the builders of civic and democratic AI systems use our data to test whether their systems are in line with best practices regarding fairness. Civic AI builders must prove their democratic principles in their actual technical specifications, not just their intentions. We will help them do just that, and release open data and open libraries, tools, and platforms so others can test their systems for operational fairness.



Repository:

<https://github.com/eticas-tech/ngi-search-eticas>

Team:

Matthew Stempeck

Country: Portugal



Contact Email:

info@civictech.guide

Other Info:

<https://civictech.guide/auditing-civic-ai/>

Fediverse Discovery Providers

Exploring decentralized search and discovery for the Fediverse

This project explores the possibilities for better search and discovery on the Fediverse in the form of an optional, pluggable service. This service should be decentralized, independent of any one specific Fediverse service and respect user choice and privacy.



Repository:

https://github.com/mastodon/fediverse_auxiliary_service_provider_specifications

Team:

Mastodon

Country: Germany, France,



Contact Email:

hello@joinmastodon.org

Other Info:

Project website with detailed description: <https://www.fediscovery.org>

FörderFunke

The matching concept behind FörderFunke turns searching conceptually on its head.

A person's profile filters various information spaces and "attracts" relevant things.

You don't have to know a benefit exists.

You don't have to search for it.

You don't have to check your eligibility.

Instead, relevant information finds you automatically.

The team focuses on state benefits for now, but it wants to expand its scope beyond that, relying on Open Source, Privacy by Design and Linked Open Data.



Repository:

<https://github.com/Citizen-Knowledge-Graph>

Team:

Country: Germany



Contact Email:

info@foerderfunke.de

HeReFaNMi

HeReFaNMi (Health-Related Fake News Mitigation) tackles the spread of health-related fake news

HeReFaNMi (Health-Related Fake News Monitoring) stands at the forefront of tackling the propagation of fake news within health-related information and restoring trustworthiness to the internet community. This pioneering project is driven by recognizing the severe consequences of misinformation, including the promotion of ineffective remedies, the downplaying of risks, and the erosion of public confidence in scientific research.



Repository:

<https://github.com/alessandrobruno10/heretanmi>

Team:

IULM University, Dr Alessandro Bruno.
University of Orleans, Prof Aladine Chetouani. ISASI CNR, Dr Pier Luigi Mazzeo

Country: Italy, France



Contact Email:

alessandro.bruno@iulm.it

InDi

Improve Privacy and Trust in Non-Personalized Information Discovery

InDi will develop an open-source search engine with key principles such as explainability, human-in-the-loop, anonymity, fairness, and inclusiveness. It will involve a community of voluntary reviewers to manually assess AI re-rankings. A blockchain ensures fair, anonymous reviews, promoting participation from discriminated minorities. Top reviewers receive crypto-assets, initially with moral value, potentially gaining extrinsic value, akin to Bitcoin's evolution.



Repository:

<https://github.com/ngi-indi>

Team:

University of Cagliari and R2M Solution

Country: Italy



Contact Email:

diego.reforgiato@unica.it

IoT twinning for digital product passports

The project, supported by the Polytechnic University of Catalonia and the eReuse circular economy community, aims to provide a search and discovery service to access public information about digital devices based on digital twins linked to physical devices through unique decentralised identifiers. The digital twins provide identifiers, details, and verifiability information as a digital product passport to enable circular management of ICT devices. The service collects and discovers points that accumulate over the lifespan of the devices, contributed by various stakeholders, to increase confidence in the second-hand device market and calculate social and environmental impacts.

**Repository:**

<https://gitlab.com/dsg-upc/ereuse-dpp>

Team:

UPC

Country: Spain**Contact Email:**

info@ereuse.org

LabDiscoveryEngine

Open Source tools to publish, share, discover and access physical educational labs and equipment

The LDE project creates an OS framework to transform physical labs into accessible online remote labs, helping share physical resources online. This initiative, a collaboration among LabsLand, H-BRS, and CPNU, aims to make labs more accessible and discoverable, particularly benefiting educational institutions. Building on WebLab-Deusto RLMS, LDE enhances cross-institutional sharing and discovery of remote labs, democratizing access to educational resources and equipment.



Repository:

<https://github.com/labsland/labdiscoveryengine>

Team:

LabsLand, H-BRS, CPNU

Country: Spain, Germany, Ukraine



Contact Email:

labdiscoveryengine@labsland.com

Other Info:

Project's website: <https://labdiscoveryengine.labsland.com/>

Ma Dada

Ma Dada aims to improve the way citizens search for and request information from public authorities.

Grounded in the French Freedom Of Information (FOI) law, Ma Dada's platform guides citizens to file requests, but also acts as an open data archive and platform for right-to-know or transparency campaigns, by publishing the whole process : the request history, the resulting correspondence, and the data obtained through it.



Contact Email:

contact@madada.fr

Repository:

<https://gitlab.com/madada-team/dada-core/>

Team:

Open Knowledge France

Country: France



Ma Dada UX

Ma Dada aims to improve the way citizens search for and request information from public authorities.

In this round, the MaDada platform, that makes more information from French public authorities discoverable to the citizens, will introduce technical improvements, such as a one-click request submission template and enhanced request visualization, to further streamline the user experience.



Repository:

<https://gitlab.com/madada-team/dada-core/>

Team:

Open Knowledge France

Country: France



Contact Email:

contact@madada.fr

MindBugs Discovery

Visual search engine and analytics for disinformation across Europe

MindBugs Discovery is a visual search engine enhancing investigations by understanding query contexts, enabling journalists and researchers to find debunked claims efficiently, saving time for detailed reporting. It offers geographical and timeline visualizations, showing disinformation trends across regions. The interactive 3D graph feature explores data dynamically, revealing patterns and connections, transforming complex analyses into actionable insights.



Contact Email:
cheresioana@gmail.com

Repository:

https://github.com/cheresioana/kg_repo

Team:

Ioana Cheres, Mihai Topor, Antonia Niculae,
Codruta Simina, Razvan Munteanu

Country: Romania



MOUSSE (Metadata fOcUsed Semantic Search Engine)

Metadata focused Semantic Search Engine

Metadata fOcUsed Semantic Search Engine is a semantic search tool powered by advanced large language models (LLMs). It is designed to enhance data discovery by focusing on metadata, making it particularly suitable for large, diverse datasets that lack structured ontologies.



Repository:

<https://github.com/LIBRA-AI-Tech/mousse>

Team:

Country: Spain



Contact Email:

hello@libraai.tech

Neural Datafari

Neural Datafari, an end-to-end gen AI search solution, with multisources crawling capabilities.

Adding RAG and vector search to Datafari for a neural end to end search solution, providing heterogeneous crawling capabilities up to a search user interface, empowering citizens and organisations with a privacy preserving complete search solution. Powered by Apache ManifoldCF and Apache Solr.



Repository:

<https://github.com/francelabs/datafari>

Team:

France Labs

Country: France, United Kingdom



Contact Email:

info@francelabs.com

OKLLM (Online Knowledge Search for LLM innovations)

Online Knowledge Search for LLM innovations

The increasing availability of LLMs is opening up a new era in online content discovery. However, LLMs are computationally intensive. Worse still, the elements generated by the AI may contain biases or false information.

As part of this research, the project aims to fill these gaps through knowledge distillation, knowledge graph generation and verification, bias detection and transfer learning.



Repository:

<https://github.com/innovationAppliedInsight/okllm>

Team:

University of Galway

Country: Ireland



Contact Email:

huan.chen@universityofgalway.ie

On My Disk: search integration

On My Disk offers personal cloud storage on private devices with secure access to third parties. PeARS is a local search engine that indexes and searches web pages while preserving privacy. The two platforms are partnering to offer a decentralised search system over personal cloud data, allowing users to search documents across networked devices. The joint project aims to create a new ecosystem that combines private hosting and distributed web search.

**Repository:**

<https://forge.bineon.team/onmydisk/onmydisk/-/tree/ngi-search>

Team:

Country: Slovenia

**Contact Email:**

support@onmydisk.com

Other Info:

<https://forge.bineon.team/onmydisk/onmydisk/-/tree/ngi-search>

<https://github.com/PeARSearch/PeARS-lite/tree/ngi-search>

On My Disk: Web UI

In this second NGI Search project, the team main milestone has been the development of a user-friendly website hosting feature that allows non-technical users to easily host static sites from their personal cloud. The combination of the On My Disk private cloud solution with the PeARS search engine offers each individual a way to store their content on their private devices, and to make it available to others searching information across the entire network. These efforts should foster more individuals and organisations to design the search engine they need for their community and run it themselves.

**Repository:**

<https://forge.bineon.team/onmydisk/onmydisk/-/tree/stable/>

Team:

Country: Slovenia

**Contact Email:**

support@onmydisk.com

Other Info:

<https://forge.bineon.team/onmydisk/onmydisk/-/tree/stable/>

<https://github.com/PeARSearch/PeARS-OMD/tree/main>

Open Data Deep Search

ODDS is an AI based search engine for open data - bringing out the most of any open data portal.

ODDS (Open Data Deep Search) is an AI based tool that acts as a data analyst:

- it finds the correct and relevant datasets
- understands the data format and structure
- queries the data to find the exact information requested
- and provides clear answers to the user's questions.



Repository:

<https://github.com/whiletrue-industries/odds>

Team:

Adam Kariv, While True Industries

Country: Israel



Contact Email:

adam@whiletrue.industries

Other Info:

<https://whiletrue.industries/odds/>

Open Politics

The Open Politics web application enables interaction with political data, from geospatial analysis to AI-guided decomposition of key issues, actors and interests and retrieval of relevant articles.

The database is reinforced by qualitative methods (e.g., tool-equipped AI chains) designed to enable the data to answer the most complex questions in the field of political science.



Repository:

Team:

Country:

Contact Email:

contact@open-politics.org

OSCAR4US (Open Source Scalable and Continuous Augmented Reality Services for Urban Spaces)

OSCAR4US towards open and interoperable city-scale, context-aware, shared AR experiences.

The OSCAR4US project, coordinated by Open AR Cloud Europe gUG, developed and advanced the Open Spatial Computing Platform (OSCP) to support city-scale location-based Augmented reality experiences. OSCP was extended in three areas: (1) OpenVPS, an open visual positioning service, (2) continuous, context-aware discovery of services and content, and (3) a new Reality Modeling Language (RML), a machine-readable description of the physical environment.



Repository:

<https://github.com/openArCloud>

Team:

Led by Open AR Cloud Europe gUG,
project partner Institute

Country: Germany, Norway, Hungary, Switzerland



Contact Email:

communication@openarcloud.org

Other Info:

For more information regarding the project deliverables, visit the project website:

<https://www.oscar4us.eu/>

Open Search Ecosystem for Technology

Towards a Linked Open Data Powered Hardware Design Search Ecosystem

OSE is leveraging ontology tools to enable editing of distributed standards, such as "Open Know-How". It is build as a stack of tools for a distributed way to index and search hardware. This allows people to find, research and combine existing technology, and make its meta-data available for powerful search applications based on Linked Open Data.



Repository:

<https://github.com/iop-alliance/OpenKnowHow/>

Team:

Open Source Ecology Germany e.V.

Country: Germany



Contact Email:

info@opensourceecology.de

QLever (SPARQL 1.1 Compliance)

SPARQL 1.1 Compliance

QLever is a highly efficient graph database that can handle RDF datasets of hundreds of billions of triples on a standard PC. It also shines on smaller datasets, processing queries faster than other engines.

NGI Search will allow an even wider usage of QLever through its compliance with the SPARQL 1.1 standard, while maintaining its high efficiency.

**Repository:**

<https://github.com/ad-freiburg/qllever>

Team:

Country: Germany

**Contact Email:**

simon.schorlepp@cs.uni-freiburg.de

Science Checker - reloaded

Help making sense about scientific facts

Opscidia has created an automatic fact-checking tool named "Science Checker". The Science Checker helps non experts to assess whether a claim is backed or contradicted by the scientific literature. The current project aims at developing a new search technology for improving Science Checker. This new search technology will allow to query the system with a simple question in natural language, it will identify the most relevant articles and extract the answer to the user's question.



Repository:

<https://github.com/opscidia/science-checker>

Team:

Opscidia

Country: France



Contact Email:

sylvain.massip@opscidia.com

SCION Browser

SCION Browser enables path-aware retrieval of web resources over the SCION Internet architecture.

SCION Browser enables path-aware retrieval of web resources over the SCION Internet architecture.



Repository:

<https://github.com/netsys-lab/scion-browser>

Team:

OVGU Magdeburg

Country: Germany



Contact Email:

simon.scherrer@inet.zurich

Search-a-licious

An open, extendable, pluggable search platform for Open Food Facts (and other open inventories).

Search-a-licious transforms large data collections into searchable content, enabling users to quickly find what they need. With powerful queries, facet exploration, and visualizations. Originally developed for Open Food Facts, it's ideal for exposing data, decision support, exploration and more, unlocking the full value of open data collections. Developers can build apps rapidly thanks to its easy configuration, reusable components and robust architecture.



Repository:

<https://github.com/openfoodfacts/search-a-licious/>

Team:

Open Food Facts

Country: France



Contact Email:

contact@openfoodfacts.org

Other Info:

<https://world.openfoodfacts.org>

Spare Cores

Spare Cores, a Python-based open-source ecosystem, offers standardized inventory and performance evaluations for compute resources of public cloud and server providers, and helps with selecting and actually starting the optimal instance type for containerized tasks, such as ML model training, ETL processes, rendering videos, traditional CI/CD tasks for software developers, or even microservices with varying compute requirements.

**Repository:**

<https://github.com/spareCores>

Team:

Country: Hungary

**Contact Email:**

social@sparecores.com

SVP64 Power ISA Vector Optimisation for Search

Power ISA Vector Optimisation for Search

This project is about hardware optimisation (specifically ISA) for common search algorithms.

It uses Libre-SOC source code, a combination of GPL and LGPL that is matching the FOSS values promoted by NGI.

SVP64 contributes to energy efficiency and environmental sustainability.



Repository:

<https://git.libre-soc.org/?p=openpower-isa.git;a=summary>

Team:

VectorCamp, Red Semiconductor, VanTosh

Country: Netherlands, Greece, United Kingdom



Contact Email:

djac@redsemiconductor.com

SWH Scanner

Harnessing cloud resources to optimize compute-heavy batch and service tasks efficiency and costs.

SWH Scanner is a free software edited by the Software Heritage Foundation (Software Heritage is under the responsibility of INRIA, the French national institute for Research in computer science and control). It is currently a multi-platform CLI that scans a source code project to discover files and directories existing in the Software Heritage archive and output a summary of the results to different formats. In the near future, SWH Scanner should become a solution to identify open source content provenance and identify potential security vulnerabilities, licensing issues, or outdated components in the software being developed or used.

The project is entirely open source, like the whole Software Heritage archive platform. The swh-scanner project is released under GNU GENERAL PUBLIC LICENSE (GPL).



Repository:

<https://github.com/SoftwareHeritage/swh-scanner>

Team:

Attila Nagy & Balazs Hodobay & Gergely Daroczi

Country: France



Contact Email:

contact@softwareheritage.org

Other Info:

<https://docs.softwareheritage.org/devel/swh-scanner/>

TalentLayer

The API layer connecting the world's

TalentLayer is an API that connects the world's marketplaces. When one platform has too much liquidity for, for example, Python Development jobs, and not enough developers to work on them, they can push those jobs to other marketplaces to ensure high fulfillment rates. This helps platforms retain users and earn on "joint transactions" between their users and users of other marketplaces. Users get don't have to use multiple platforms to access work, and have one reputation across platforms.



Repository:

<https://github.com/TalentLayer>

Team:

Kirsten Pomales, Romain Martin, Quentin de Chateaubriant

Country: Spain, Portugal, France



Contact Email:

kirsten@talentlayer.org

Other Info:

<https://talentlayer.org>

<https://docs.talentlayer.org>

TEThYS (Topics Evolution That You See)

Building the next-generation open-source Web topic explorer for big datasets of textual documents

TETYS provides improved access to technical/general text to quickly grasp their topics' temporal trends. It is composed of a pipeline for ingesting huge data corpora, extracting relevant topics clustered along orthogonal dimensions, and an interactive dashboard, supporting visualization of topic word clouds and exploration of their temporal series. TETYS aims to make Web users

- 1) aware of the semantic content of text datasets
- 2) understand temporal evolution of the topics.



Repository:

<https://github.com/FrInve/TETYS/>

Team:

Anna Bernasconi (project lead), Francesco Invernici, Stefano Ceri (Politecnico di Milano)

Country: Italy



Contact Email:

anna.bernasconi@polimi.it

Other Info:

<http://gmql.eu/tetys>

<http://gmql.eu/cortoviz/>

The World Literature Knowledge Graph

The World Literature Knowledge Graph

WL-KG is conceived as an open community where:

1. sources from different digital archives are harmonized and referenced by using a common semantic model, in order to assess to which extent they adequately represent non-Western writers;
2. Information Extraction (IE) methods are implemented to gain knowledge from unstructured information about authors, works, and contents produced by WL-KG users.



Repository:

<https://github.com/world-literature-kg>

Team:

University of Turin (M.A. Stranisci, R. Damiano, V. Patti), University of Bari (S. Ferilli, E. Bernasconi)

Country: Italy



Contact Email:

rossana.damiano@unito.it

Trallie

Turn free-form descriptions into tables

Trallie is a framework that leverages the power of large language models (LLMs) to reimagine information extraction (IE) with or without user guidelines. Instead of relying on labelled examples or copious amounts of training on your data collection, Trallie requires very few or no representative examples of the data to convert into a structured format. Trallie operates with three core objectives:

- Minimal manual annotation
- Flexibility with document formats and languages
- Accessibility to non-expert users.



Repository:

<https://github.com/PiSchool/trallie>

Team:

Pi School

Country: Italy



Contact Email:

info@pischool.com

Other Info:

<https://picampus-school.com>

Trust4AI (Trustable AI-Driven Internet Search)

Trustable AI-Driven Internet Search

Large Language Models (LLMs) like ChatGPT are changing how we search and decide online. Integrated into Bing, they show promise but raise safety and trust concerns. The EU AI Act aims to regulate AI, demanding trustworthiness from companies. Currently, ensuring AI compliance is manual, slow, and prone to errors. This project will investigate automating compliance checks using metamorphic testing to make the process more efficient and reliable.



Repository:

<https://github.com/orgs/Trust4AI/repositories>

Team:

Aitor Arrieta (Mondragon Univeristy),
Sergio Segura (University of Seville)

Country: Spain



Contact Email:

aarrieta@mondragon.edu

Other Info:

Project link with all information: <https://trust4ai.github.io/trust4ai/>

TrustSearch

Promote critical thinking by providing news articles of difference stance and political orientation

Trust in information is at a record low. The goal of TrustSearch aims is to promote critical thinking offering users information about the stance of articles towards important topics and to encourage them to consume more pluralistic information from media of different political orientations. The tool implements different use-cases to detect an article's stance on some controversial topics such as climate change, immigration, and vaccination against COVID-19.



Repository:

<https://github.com/nikolamilosevic86/verif.ai>

Team:

Universitat Politècnica de Valencia (Paolo Rosso), Athens Technology Center (Danae Tsabouraki)

Country: Spain, Greece



Contact Email:

proso@dsic.upv.es

Other Info:

Project info: <https://www.prhlt.upv.es/trustsearch/>

Demo: <https://trust-search.lab.atc.gr/>

Paper: https://ceur-ws.org/Vol-3729/d8_rev.pdf

Verif.ai

No more searches, just verifiable answers!

Verif.ai project aims to address problem of hallucinations in generative large language models and generative search. In the core of the engine is generative search engine, powered by open technologies. However, generative models may hallucinate, and therefore Verif.ai is developing a second model that would check the sources of generative model and flag any misinformation or misinterpretations of source documents.



Repository:

<https://github.com/nikolamilosevic86/verif.ai>

Team:

Nikola Milosevic, Lorenzo
Cassano (Bayer), Adela Ljajic, Milos
Kosprdic, Bojana Basaragin (SIAI)

Country: Germany, Serbia



Contact Email:

nikola.milosevic@bayer.com

Other Info:

Follow our project development updates at: <https://www.linkedin.com/company/sct-lab>

Visio Creativia

Bridging Realms Through Digital Artistry

Originally called MetaVision, this project provides decentralized solutions, including blockchain and distributed ledger technologies.



Repository:

<https://git.wur.nl/spyro004/sct-project>

Team:

Wageningen University and Research

Country: Netherlands



Contact Email:

will.hurst@wur.nl

WAISE – Wiki Artificial Intelligence Search Engine

AI-driven, permission-aware chatbot across XWiki and apps

WAISE is an open-source conversational AI server built on XWiki. It uses a vector-based RAG engine and works with OpenAI-compatible LLMs (such as LocalAI) to answer user questions based on indexed content, while always respecting permissions. It is an XWiki extension that supports external data ingestion via a REST API and embeds securely in other applications.



Repository:

<https://github.com/xwiki-contrib/ai-llm>

Team:

XWiki SAS, project lead Michael Hamann (with contributions from Paul Panțiru, Vincent Massol, Ludovic Dubost and others)

Country: France, Romania



Contact Email:

contact@xwiki.com

Other Info:

WAISE is part of the NGI Search initiative and received EU funding that allowed XWiki to dedicate full-time development resources to build a privacy-preserving, permission-aware AI search assistant. You can learn more about the project from our detailed blog post "Leveraging AI to enhance open-source collaboration with WAISE" (<https://xwiki.com/en/Blog/Leveraging-AI-to-enhance-knowledge-sharing-with-WAISE>), which covers architecture, integration with XWiki and external apps, and how NGI support accelerated core features. Our earlier post "The first AI assistant in XWiki is here! (BETA)" (<https://xwiki.com/en/Blog/AI-assistant-in-XWiki/>) covers the beta release of our AI assistant. It explains what makes it different from other language models and explores key use cases like summarization and multilingual support. The post also outlines the challenges we're tackling around privacy, bias, and energy consumption.

For a deeper technical dive into the architecture, installation steps, and our future plans, check out our design documentation: <https://design.xwiki.org/xwiki/bin/view/Proposal/X-AI/WAISE>.

WASPer

Synthetic propaganda detection in the age of generative AI.

With the leaps in generative AI and the ease of access to it, we see an expected rise in its utilization as part of cognitive warfare efforts. Coordinated communication attacks are a segue to state capture and they have become easier to design. The WASPer series of models and taxonomies helps media and communication professionals both in the public and private sectors, by analysing whether content is synthetic, whether it contains propaganda, and if it does - what type of propaganda it is.



Contact Email:
info@identrics.ai

Repository:

Team:

Yolina Petrova, Boryana Kostadinova,
Bogomil Katanov, Nikola Blajev

Country: Bulgaria

