Istat | Istituto Nazionale di Statistica

Geneva, 22 November 2023

HLG-MOS Workshop on the Modernisation of Official Statistics

Can AI better satisfy users of statistical information? A case study in Istat

MAURO BRUNO - Istat - Directorate for Methodology and Statistical Process Design
SARA LETARDI - Istat - Directorate for Information Technologies
MARCO SILIPO - Istat - Directorate for Information Technologies
MICHELA TROIA - Istat - Directorate for Communication

Outline

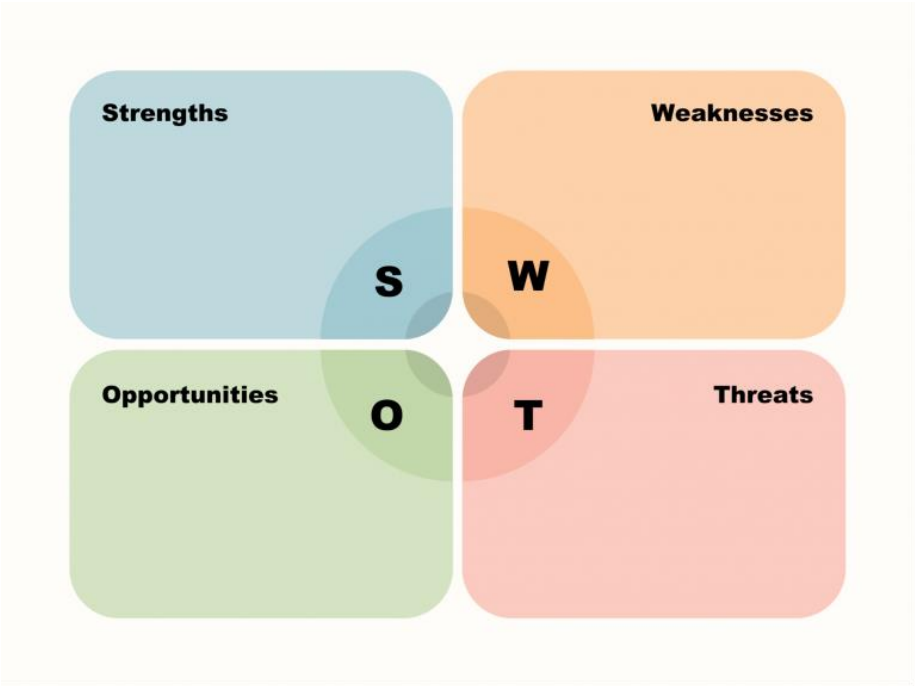
- Introduction
- A SWOT analysis (Strengths/Weeknesses-Opportunities/Threats)
- Istat use case
- Pilot research projects
- Chat Bot Project
- Benefits and Constraints of an AI solution in Istat
- Conclusions and future perspectives

Introduction

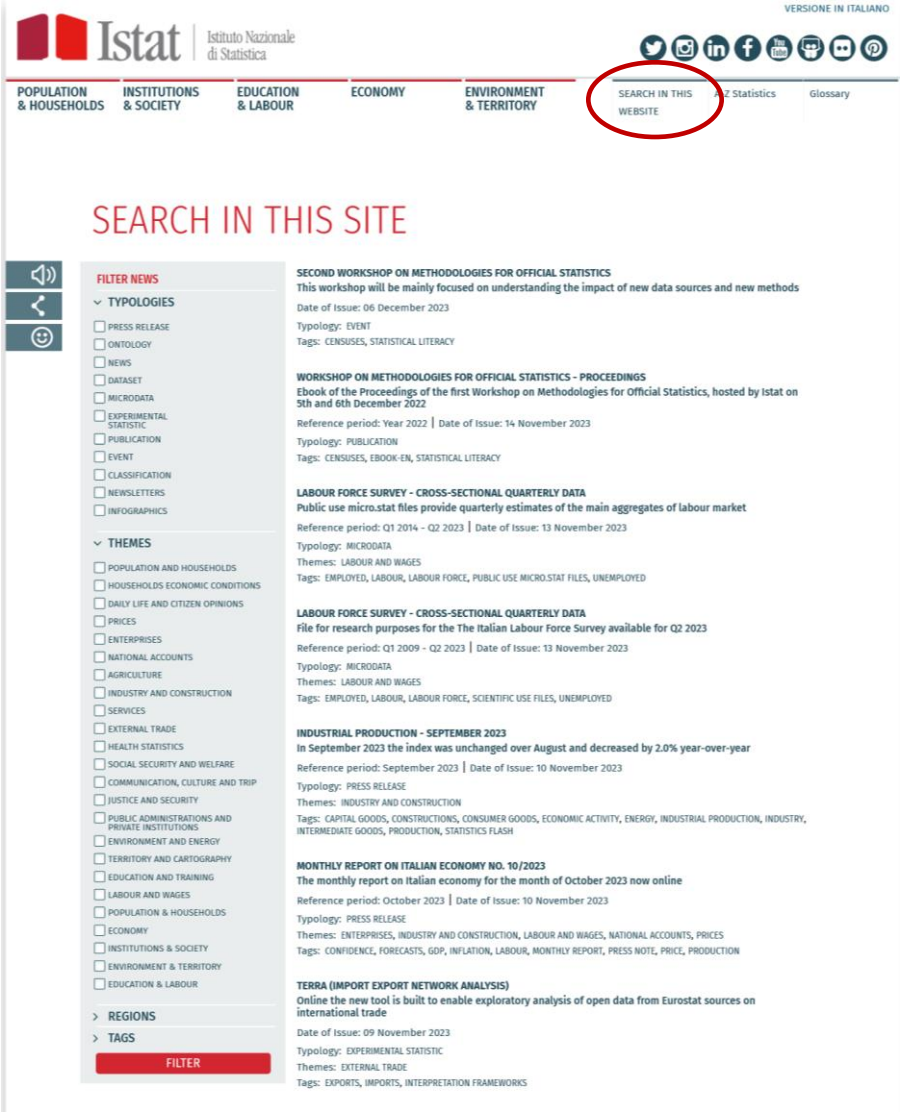
- Generative Artificial Intelligence (AI) offers a wide range of capabilities, in various field. Its capabilities continue to evolve and expand as technology advances, making it a key component of the future of AI
- It is important to understand the implications and possible applications of generative AI in official statistics, **to make the most of its advantages and ensure its responsible use**
- In this presentation we would like to share our experience, illustrating:
  - the scenario analysed,
  - the choices made,
  - the steps taken
  - and those still to be taken

Internal factors: Strengths

- Vast wealth of information
- Istat is the most important producer of official statistics at national level
- Essential active role in statistical literacy

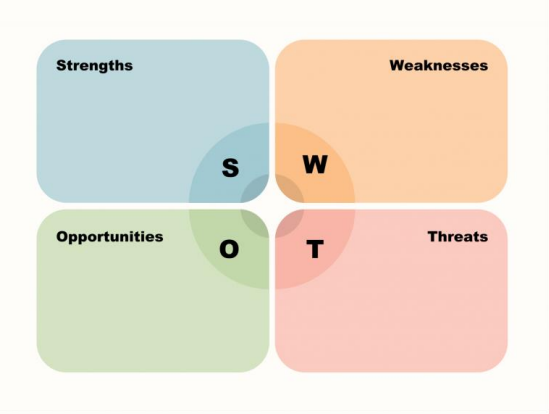


Internal factors: Weeknesses



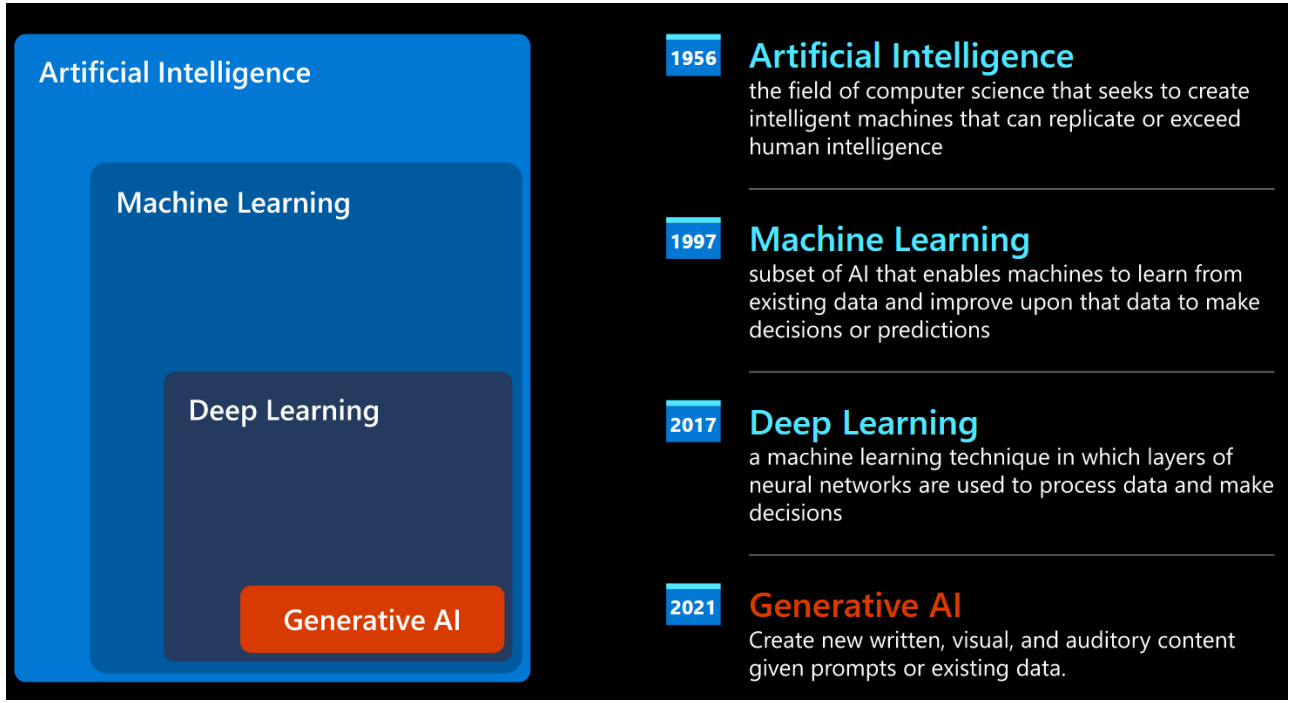
The current search on the institutional site is not effective:

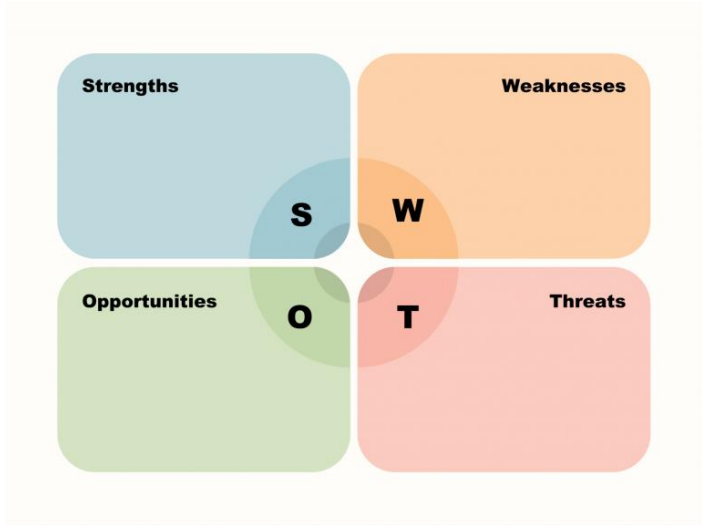
- Full-text searches are not possible
- Available filters are grouped in long lists
- The results obtained cannot be sorted by relevance



Extenal factors: Opportunities

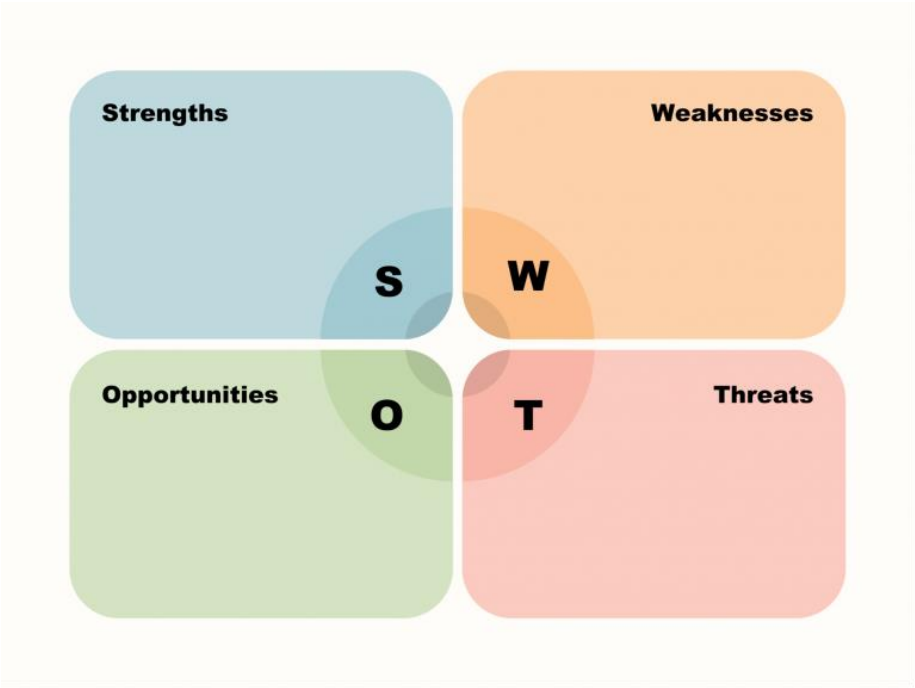
Incorporating AI-based chat on research organization's website can provide several benefits, enhancing user experience and supporting the organization's mission





Extenal factors: Threats

- High risk of users leaving the site (without a effective search engine)
- Misinformation
- Fake news



Istat use case

Experimentation **aimed at improving user experience on Istat institutional website through the use of AI** (semantic search and generative AI):

- to provide a more effective and efficient service to users accessing official statistical information
- to take the opportunity offered by AI to provide high quality data to a wider audience

This project requires **communication**, **technological**, and **methodological** skills

Istat use case

Challenge

Starting from the sound experience of searching our data on the institutional site, we would like to offer the user the possibility to perform **traditional searches** and **semantic searches**, i.e., natural language query, and return the results in a 'generative' way.

The challenge is therefore to propose **a dialogue between a user and a virtual assistant.**

Output

A support system capable of answering natural language queries based on a **‘specific’** context (**Istat documents**), but with the possibility of integrating **‘general’** information (**from GPT models**).

Pilot research projects

1) Development of a **semantic search component** that allows users to browse the website, with all its information assets, using natural language

2) Integration of a **chatbot** to provide detailed and relevant answers, based on the content of Istat's website

3) Implementation of an open-source **chatbot** (we have a prepared a surprise)

We will focus on the chatbot projects!

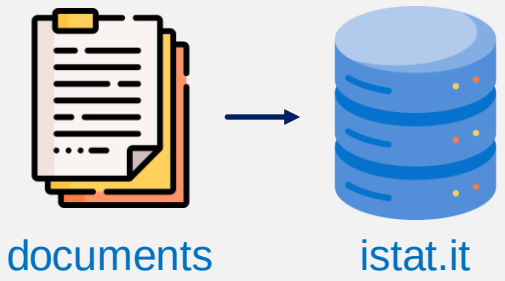
Benefits of an AI solution in Istat

- Subset of **specific and updated documents** to search on versus an external search engine that searches on everything
- Ability to minimize the frequency of **hallucinations** compared to an external chat
- Ability to make a model that returns an answer upon **verification that the given source actually exists**
- The search engine will also **search attachments** and not just HTML pages, thus increasing the amount of information that could potentially be released
- Multiple possibilities for displaying results (tables, diagrams, csv/xlsx reports, etc.)
- **…**

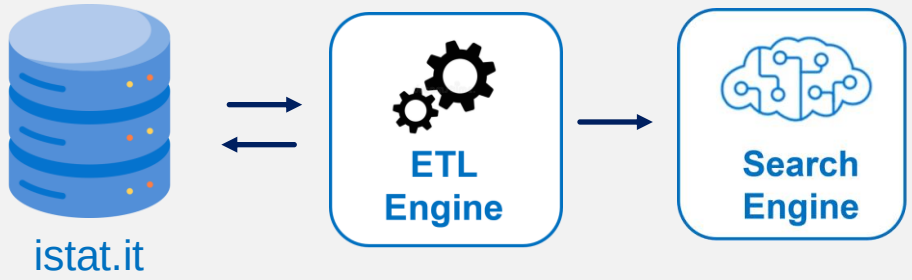
Constraints for an AI solution in Istat

- Provide answers based only on **Istat content** and not the entire web
- Provide answers with relevant data source (**link to specific document**)
- Do not make predictions and possibly answer that the data does not exist
- Filter content effectively and individually, setting different **levels of severity** - low, medium, or high - to prevent dissemination of inappropriate messages

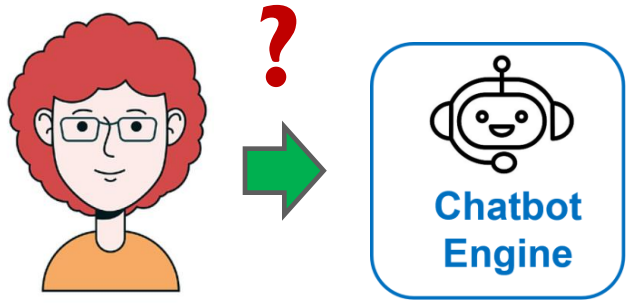
Chatbot in action



1. Loading relevant documents in Istat's archive



2. The **ETL Engine** detects the newly uploaded documents and transforms them to load them in the **Search Engine** (vector data base)

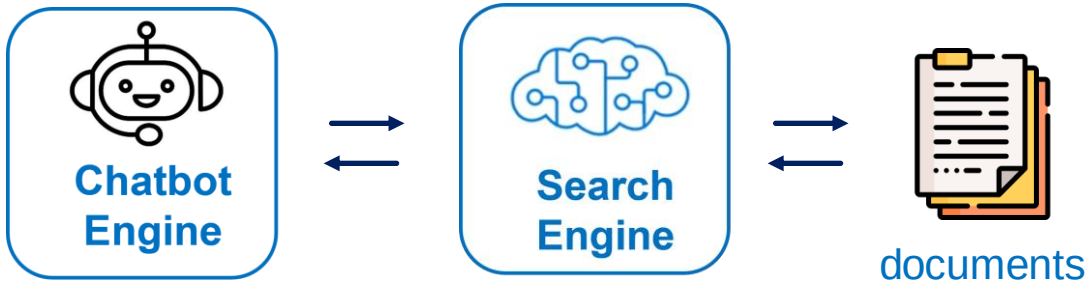


3. The user queries the **Chatbot Engine** about the documents in Istat's archive

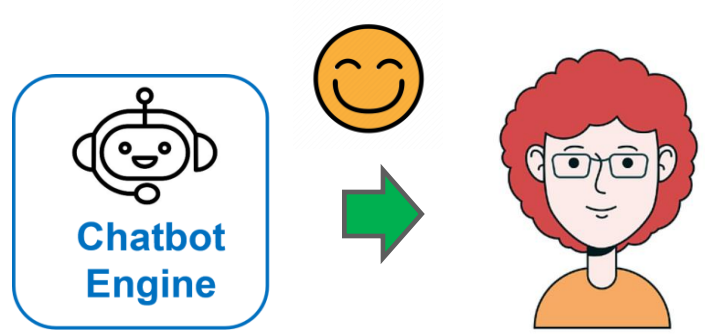
Istat - knowledge base

Istat

Chatbot in action

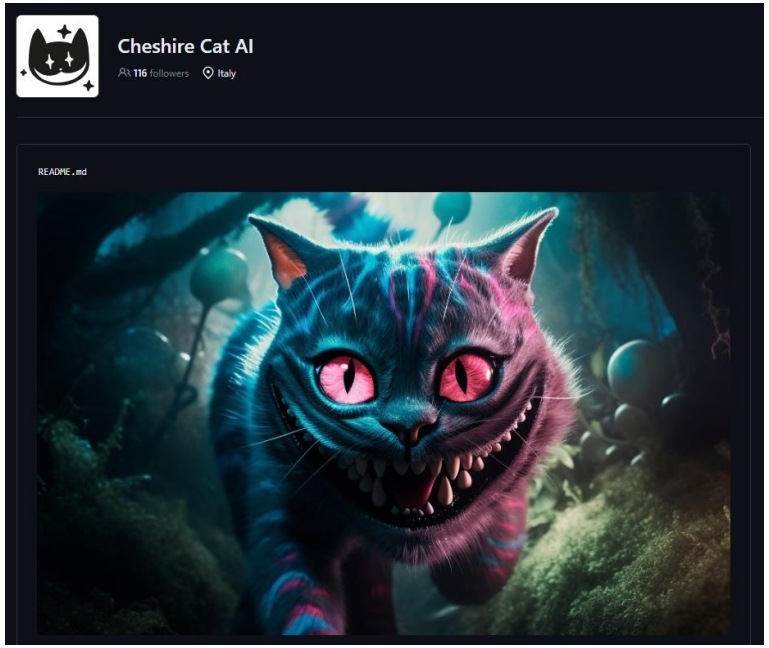


4. The **Search Engine** retrieves the most relevant documents related to the user's question and provides them to the **Chatbot Engine**



5. The **Chatbot Engine** provides a relevant response to the user

AILab: implementation of an open-source **chatbot**



Open source project: Cheshire Cat AI

We started working on a chatbot, based on the Cheshire Cat AI project, designed to **deliver responses customized to a collection of documents fed into the system**

Source: DALL·E

Marco Silipo will provide a wonderful demo on our chatbot

Conclusions and future perspectives

Collect

Use a chatbot for survey questionnaire filling support, increasing the quality of data collected and the efficiency of data collection processes

Disseminate

1. Semantic search on the site
2. Semantic search for SDMX data base
3. Contact center for user support (e.g., searching data and indicators)

AILAB

Creation of a **research AI Lab** to:

- support the growhing of IT capabilities and competencies in Istat in AI
- increase digital skills in AI among users by promoting AI-aware use

Istat

Thank you!

MICHELA TROIA | mchela.troia@istat.it

SARA LETARDI | sara.letardi@istat.it

MAURO BRUNO | mauro.bruno@istat.it

MARCO SILIPO | marco.silipo@istat.it

Istat | Istituto Nazionale di Statistica