

Artificial intelligence as a support for survey respondents: defining the process of Istat new AI service

Paola Bosso¹, Gabriella Fazzi¹, Paolo Francescangeli¹

¹ISTAT Italian National Institute of Statistics

Abstract

Currently, responding units involved in the surveys of the Italian National Institute of Statistics (Istat) can request assistance and support in accessing and navigating the data acquisition systems, as well as with legal obligations or doubts about the survey's content. Assistance service is provided through synchronous (toll-free number) and asynchronous channels (dedicated email). The service is managed by a specialized Contact Center acquired from an external supplier.

The current management of service requests is exclusively interactive and is based on two levels: first-level assistance is provided by operators from the external company (Contact Center assistance) who solve the most common problems using FAQs. The second level of assistance focuses on cases with a higher degree of complexity and is provided by Istat experts (Istat assistance).

In 2023, 260.000 tickets were managed by phone and email to assist households and enterprises, and were resolved at either the first or second level of assistance. In 2024, Istat plans to introduce a new integrated assistance service, with the objectives of optimizing resources and simplifying communication with Istat.

The new service provides automatic interaction processes with respondents, also through the use of Artificial Intelligence solutions, in a multi-channel perspective (telephone, email, pec, web, webchat, social media). Automation is achieved not only through integrated management of different systems and data flows, but also through the use of machine learning algorithms to process and respond to respondent requests. Experimental analyzes on tickets acquired over time show that a significant percentage of them concern issues that can be managed using Artificial Intelligence (AI). The use of AI can help reduce waiting times for respondents and costs for Istat.

This paper describes an automated procedure for classifying assistance requests and responding to the most frequent ones by using natural language processing techniques and supervised classification algorithms. Therefore, an initial level of automated assistance is introduced, and a new process is designed to direct requests to the appropriate level of assistance: AI assistance, Contact Center assistance, or Istat expert assistance. Additionally, we address privacy concerns that may limit certain results and applications of AI.

Keywords: User Assistance, Artificial Intelligence, Process Architecture, Classification

1. Introduction

Since 2016, Istat has used a specialized Contact Center service to support the units involved in statistical surveys. Supporting 80 surveys, including sample and census conducted with CAWI (Computer Assisted Web Interviewing) mode, it is a key component of the centralization

strategies of Istat's data collection processes, launched in 2016 as part of the modernization process of the Institute. Centralized assistance improves the efficiency of data collection processes, optimizes human resource allocation and reduces the burden on respondents. The management of the service is entrusted to specialized external operators, chosen through periodic public procedures, which provide assistance through their human and technological resources.

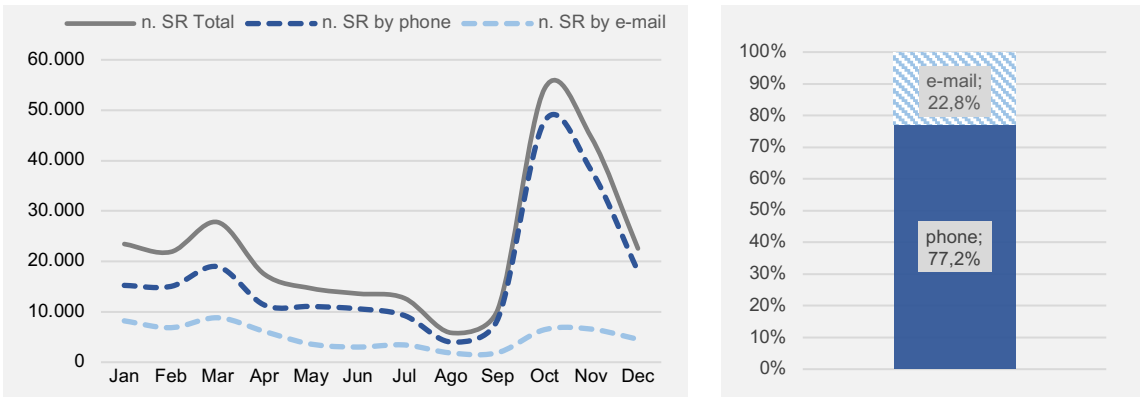
2. The assistance process up to 2023

Statistical units access the assistance service through a toll-free number (synchronous channel), dedicated email or certified electronic mail (asynchronous channels). The management process's assistance request operates with two levels of specialization:

- *First level:* external operators identify statistic units from assistance requests, classify the request, attempt to resolve recurring and simple issues, escalate complex issues to Istat experts;
- *Second level:* Istat experts are dedicated to addressing more complex requests.

In 2023, the assistance service handled 269,359 requests, totalling 1,435,047 minutes of assistance time. Request volumes varied throughout the year, with the highest peaks observed in the October-December quarter, coinciding with the population census (Fig.1). Telephone channel was the most utilized by statistical units comprising 208,013 service requests (77.6% of cases), followed by email with 61,346 service requests (22.4% of cases).

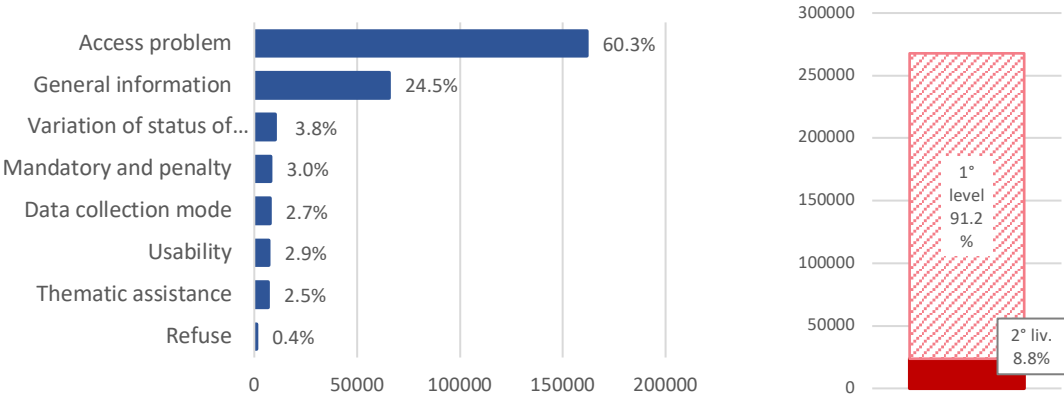
Figure 1: Requests monthly distribution in 2023 (on the left) – Channel used 2023 (on the right)



The topics of incoming service requests can be classified in eight classes, which highlight a predominant share of *access problems* and *general information* (Fig.2). The 91.2% of cases were resolved by external operators (at first-level specialization) through FAQs and

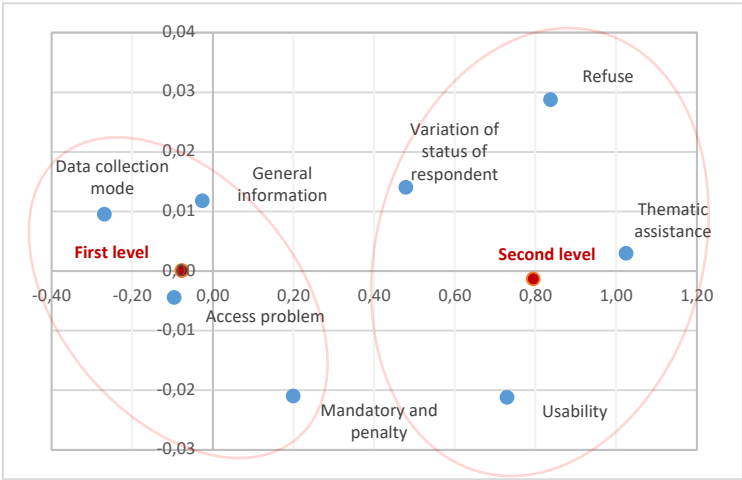
standardized procedures, while only 8.8% required intervention from Istat experts (second-level specialization).

Figure 2: Categories of request 2023 (on the left) – Resolution level 2023 (on the right)



Simple correspondence analysis of request categories by resolution levels was conducted to find out insights on standardization. The analysis highlight an association between the two dimensions: access issues and requests for general information are associated with the first level, while thematic requests, usability and refusal to cooperate require second-level intervention as they concern non-standardisable contents.

Figure 3: Correspondence Analysis of Categories and Resolution level



The 2023 data indicates that the centralized assistance service has indeed contributed to improving the productivity and efficiency of data collection processes at Istat, by delegating routine tasks to external resources. However, some weaknesses persist in the current assistance model:

- Supplier turnover, which involves long implementation, training periods and limited adaptability to Istat needs;
- Procedures that are not fully automated place the entire burden of the support process on operators, resulting in critical management of peak periods, increased unanswered tickets, longer processing times and higher error rates;
- Unclear classification methods that prevent the integration of automated process.

3. Architecture of the new Assistance Service

In 2024 Istat re-design the assistance service with the introduction of important innovations.

The features of the new service includes:

- Proprietary multichannel platform for managing assistance requests;
- Automatic approach in collecting information and routing requests;
- Artificial Intelligence solutions for the management of routine requests and activities;
- Three resolution levels with the introduction of an AI agent at first level.

Automation is achieved not only through integrated management of systems and flows different data, but also through the use of machine learning algorithms to manage respond to requests.

3.1 Proprietary Multichannel platform

Managing directly the platform allows Istat to customize functions and procedures in relation to its current and emerging needs. One of the key differences in the transition from the new to the old system is the expansion of contact channels offered to the user. Until now, in fact, to receive assistance, it was possible to call a toll-free number (there are three different toll-free numbers aimed at businesses, institutions, or households) or write an email to a specific survey address. With the new system, a single toll-free number will be active, identifying the Istat service; the email will always remain active; a chatbot channel managed initially by artificial intelligence and secondarily by a real operator will be added; and it will be possible to access the dedicated portal and open a ticket on the web. Later it is also planned to activate a support channel via WhatsApp.

it is important to note that whatever channel the user accesses, all communications will be handled in a coordinated manner. It will then be possible to keep track of requests from users who register on the web, ask a question in chat, and then call the CC, where they have identified themselves with the same ID. All personal data information needed to link the requests received is kept for 18 months after the investigation is closed and then deleted to protect the privacy of users

3.2 Request Domain Classification and patterns

The current ticket classification system used in the contact center has been defined incrementally, with new entries added as new FAQs and surveys were taken over by the center. Contact Center operators use these FAQs, which are organized into three levels, to respond to user requests. However, each survey adopts different classification systems, and it is left up to the operator to register the request by assigning the correct class in the system. This complex and non-standardized way of operating leads to several classificatory problems, including the lack of a unique "fundamentum divisionis," resulting in classes that are not mutually exclusive. This means that two identical requests may be classified in two different classes by different operators or even by the same operator at different times or for different surveys. The 24 categories are shown to the operators without a structured order but can be organized into eight classes¹.

To address these issues, a text analysis of the tickets arrived at a single survey was conducted using manual and automatic classifier systems. The result of the analysis was extended to a more extensive dataset (Bruni et al, 2023). We have also analyzed several files of FAQs related to households or business surveys to identify the more important categories of problems. All these results have been used to design a newly revised classification that is more grounded in data and useful in identifying at a glance the main problems that respondents may encounter in a survey. The newly revised classification is designed to be responsive to different channels of contact used by respondents. It has a level for IVR, where only the more general categories are proposed, a level for the operators (with more detailed categories), and a level for AI, which can manage a higher complex number of classes.

The IVR, which relies only on audio communication, asks the respondents the main reason why they are contacting the CC: they can have difficulties in accessing the portal or the questionnaire; request information on legislation, privacy, or obligation to answer; information and clarification on the survey or questionnaire; information on a fine; want to book interviews.

¹ 1) Access problems (access, status and how to submit the questionnaire; lost credentials; registration info; problems accessing the system; password reset); 2) general information (Data collection calendar; clarifications; general information on data collection mode; general information about the survey; information on questionnaire contents; method of sample selection; request for information); 3) Variation of status of respondent; 4) Usability (How to send the questionnaire; Problems with compilation - Blocking errors; Role-based access control; questionnaire navigation); 5) thematic assistance (Information on specific questionnaire variables; Identification of survey recipients); 6) Normative and fines (Regulatory aspects; Obligation and sanctions); 7) Data collection mode (Method of data collection; Interview Requests); 8) Refuses to answer.

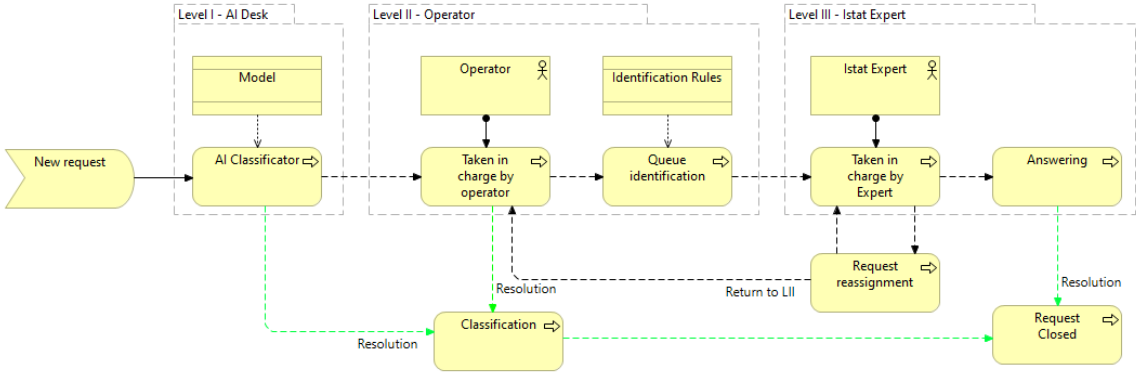
For the web, the categories are differently worded and articulated², and there is some work in progress in adopting a user-centered design approach to design a better user experience.

The AI agent (chatbot) or the CC operators can adopt the more detailed classification, based on 54 different categories that specify uniquely the motivation of the request, with less redundancy or ambiguity. A knowledge management system supports the CC operator in classifying the request into the more appropriate categories.

3.3 Process Outline

Multi-access channels collect new requests using a template with a set of minimum level of information from the sources. The request must include respondents and the desired survey identification with the actual assistance request transcription. Request flows to increasingly complex levels of assistance (fig 4).

Figure 4: Three levels Process Outline. Green path is the routing for solved requests



If it is possible to select a satisfactory article by AI or Human Operator, the request is re-directed towards the green path in the figure (4) and closed as resolved. Whatever AI fails to classify, passes on a second level of assistance, and then at the third. At the second level, Operators job is to try understanding the problem and forwarding complex problems to the third level competent thematic experts. Routine requests are usually resolved early at the first level. Automated request classifying and standardization also makes it possible to manage request from multiple channels seamlessly. Standardizations also helps manage the same classes of issues in the same way, which helps processing faster and with less error rates. Moreover, there could be issues that were not originally managed by any standard FAQ or Resolution Procedure. So Thematic experts can write new Knowledge Articles on unattended study cased that were not previously covered by any other Article or FAQ and forward them to

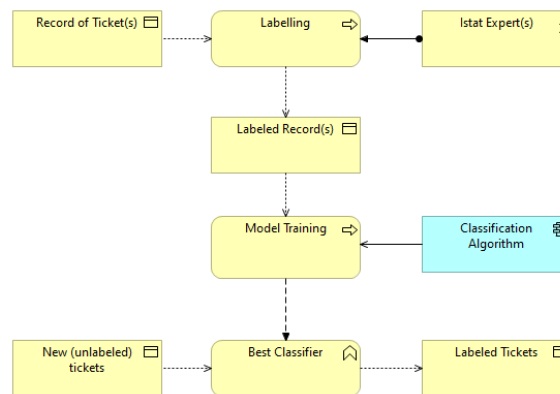
² A - Accessing and navigating the questionnaire/survey portal; B - Regulatory aspects; C - Sample composition; D - Interview booking; E - Finding and dispute report; F - Information about the survey and questionnaire contents; G - Communication with Istat; H - Other

a supervisor to approve and put it into the knowledge base for future reference. These new articles and FAQs are later processed to refine the AI model, so that new issues of the same new class can be processed at the first level of assistance.

3.4 Predictive AI for request routing

Predictive Artificial Intelligence is used to classify tickets so that they can be easily associated with the corresponding FAQ and articles in the Knowledge Base. The subsequent figure shows the details of a generic classifier process. Parameters and methods details depend on the actual task to be performed.

Figure 5: Generic classifier Process Outline



Classification Process encompasses three main phases:

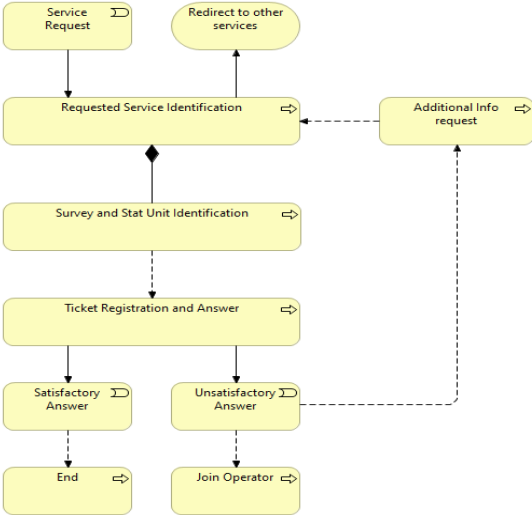
- Training Phase: Training process takes part of set of labeled records of tickets and is used to build a model. A Model is an algorithm with a set of hyperparameters to be adjusted for the purpose of classification. Human experts set the labels in the set, or these are taken from existing knowledge.
- Test Phase: Part of the labeled set is used to evaluate the classifier performance using the trained model interactively.
- Classification Phase: When the model has reached the desired precision, it will be used to classify unknown inputs.

3.5 Generative AI for Chatbot answering

Chatbot makes it possible to integrate first level assistance directly in the web channel. Just like any other input channel in the multichannel model, Chatbot first task is to make sure to select assistance requests to be forwarded to the second level of assistance and redirect any other request to their respective services. After that, it needs to identify the statistical unit and

the survey or classify the request based on its domain for further assistance by an operator as sketched in the next figure.

Figure 6: Chatbot Process Outline



With the introduction of generative AI, the system can attempt an answer which, based on the satisfaction level from the requesting user, can develop in further information requests, sustaining a full dialogue with the user. Answers are based on patterns evaluated from the classified domains of requests. Whenever the Chatbot is able to satisfy the user needs, the ticket can be closed as resolved. In other cases when answers are not satisfactory, instead it will forward the ticket to the second level of assistance with a human operator in charge.

Chatbot can access web session to get useful information such as statistical unit ID codes, language and such that need to be inputted with other assistance channels such as the telephone call. If such information is needed when user is not logged in, the Chabot itself can activate a pattern behavior with a log in request.

4. Results

To quantify the efficiency improvement of the new assistance service, we analyzed two primary indicators: case deflection and time savings. Case deflections indicates the decrease in cases that require the direct intervention of a human operator, thanks to the system's ability to manage certain incoming requests in self-service mode, via AI operators. Studies assessing the economic impact of new information technologies suggest deflection rates between 40% and 50%; we adopted Forrester's conservative assumption, setting the maximum deflection percentage at 40%. Time savings refer to the reduction in operator time required to manage requests using the automatic features of new technological solutions. Studies suggest an average reduction in ticket management time of 21%. The maximum expected levels of Case

Deflection (40%) and Time Saving (21%) will be gradually reached in the long term. Beginning with a reference assistance volume of 1,435,048 minutes in 2023, the application of successive reductions of 40% and 20% leads to a steady-state annual assistance volume of 680,213 minutes, resulting in time and cost savings of 51.6%.

References

- Bruni R., Bianchi G., & Papa P. (2023). Hyperparameter Black-Box Optimization to Improve the Automatic Classification of Support Tickets. *Algorithms*, 16(46), 1-14.
<https://doi.org/10.3390/a16010046>
- Bianchi G., Bellini G., Bosso P., Papa P. A machine learning based help-desk approach for units involved in official surveys, UNECE Expert Meeting on Statistical Data Collection 2022, Rome 26-28 October 2022.
- Bianchi G., Golini N., Righi P. Big Data and Official Statistics: some evidences. 51st Scientific Meeting of the Italian Statistical Society (SIS2022), Session Sampling techniques for big data analysis, Caserta 22-24 June 2022.
- Bianchi G., Bellini G., Di Paolo Gg., Papa P Towards a selective automation process of assistance to the survey units included in business surveys, 2022 Business Data Collection Methods Workshop, Session T1 Language processing and machine learning techniques, Statistics Norway, Oslo, 13 - 15 June 2022.
- Bruni R., Bianchi G. Website categorization: A formal approach and robustness analysis in the case of e-commerce detection. *Expert Syst. Appl.* 2019, 142, 113001
- Big Data Committee (2018) Annual Report. Italian National Institute of Statistics (Istat), 695 Rome, Italy. ISBN 978-88-458-1962-9
- Bruni R., Bianchi G., Scalfati F. Identifying e-Commerce in Enterprises by means of Text Mining and Classification algorithms. *Mathematical Problems in Engineering* Vol 705 2018, n. 7231920, 2018
- Bruni R.; Bianchi, G. Effective Classification using Binarization and Statistical Analysis. *IEEE Trans. Knowl. Data Eng.* 2015, 27, 2349–2361.