Article

Potential Impact of Using ChatGPT-3.5 in the Theoretical and Practical Multi-Level Approach to Open-Source Remote Sensing Archaeology, Preliminary Considerations

Nicodemo Abate 1,*; Francesca Visone 1,2; Maria Sileo 1; Maria Danese 1; Antonio Minervino Amodio 1; Rosa Lasaponara 3 and Nicola Masini 1

1 Consiglio Nazionale delle Ricerche, Istituto di Scienze per il Patrimonio Culturale, C.da S. Loja, sn., 85050 Tito Scalo, Italy; francesca.visone@unibas.it (F.V.); maria.sileo@cnr.it (M.S.); maria.danese@cnr.it (M.D.); antonioninervinoamodio@cnr.it (A.M.A.); nicola.masini@cnr.it (N.M.)
2 DICEM (Dipartimento delle Culture Europee e del Mediterraneo), Università degli Studi della Basilicata, Via Lanera, 20, 75100 Matera, Italy
3 Consiglio Nazionale delle Ricerche, Istituto di Metodologie per l’Analisi Ambientale, C.da S. Loja, sn., 85050 Tito Scalo, Italy; rosa.lasaponara@cnr.it
* Correspondence: nicodemoabate@cnr.it

Abstract: This study aimed to evaluate the impact of using an AI model, specifically ChatGPT-3.5, in remote sensing (RS) applied to archaeological research. It assessed the model’s abilities in several aspects, in accordance with a multi-level analysis of its usefulness: providing answers to both general and specific questions related to archaeological research; identifying and referencing the sources of information it uses; recommending appropriate tools based on the user’s desired outcome; assisting users in performing basic functions and processes in RS for archaeology (RSA); assisting users in carrying out complex processes for advanced RSA; and integrating with the tools and libraries commonly used in RSA. ChatGPT-3.5 was selected due to its availability as a free resource. The research also aimed to analyse the user’s prior skills, competencies, and language proficiency required to effectively utilise the model for achieving their research goals. Additionally, the study involved generating JavaScript code for interacting with the free Google Earth Engine tool as part of its research objectives. Use of these free tools, it was possible to demonstrate the impact that ChatGPT-3.5 can have when embedded in an archaeological RS flowchart on different levels. In particular, it was shown to be useful both for the theoretical part and for the generation of simple and complex processes and elaborations.

Keywords: ChatGPT-3.5; Google Earth Engine; Sentinel-2; archaeology; remote sensing

1. Introduction

The last three decades have been strongly marked by the impact of technologies on human life and, in particular, by their unprecedented and widespread vertical and horizontal penetration into everyday life. This impact has obviously occurred in all human activities, as well as in archaeology. With this respect, a massive technological revolution has dramatically influenced documentation techniques (i) before, (ii) during, and (iii) after excavation. Non-invasive archaeology has proven to be extremely useful in understanding or hypothesising the presence of possible remains of archaeological interest under the ground in the stages prior to archaeological excavation by providing information on the nature of the surface and subsurface using remote sensing (RS) and earth observation (EO) techniques [1–6].

RS and EO applied to archaeology are not a recent discipline. Studies in this field can be found as early as the second half of the 19th century, such as those by F. Stolze and F. C. Andreas in 1874, in Persepolis in Iran [5,7,8], and continuing with (ii) G. Boni in the Roman Forum (1899), and later (1908 approx.) in Venice, Ostia, and Pompeii [7,9,10].
A further development in RS for archaeology (RSA) occurred immediately after the Second World War [4,11–24] and subsequently with the development of satellite missions from the 1960s onwards [25–33].

The technological development of satellites and sensors has produced a major change in remote sensing and opened up new perspectives for archaeological prospecting activities [34–37]. The launch of NASA’s (National Aeronautics and Space Administration) Landsat missions represented a real change in the RS applied to the CH. In particular, in 1972, the US government distributed the data to scholars from all over the world and renamed the mission as Landsat [11,38–40].

During the 1980s, archaeologists started to structure a methodology of RSA [41–43]. During the same years, there was the First International Conference on Remote Sensing and Cartography in Archaeology and the creation of the European Remote Sensing Centre in Strasbourg, now the European Space Agency (ESA). In 1984, the first NASA-sponsored conference on RSA was held, organised by Tom Sever and James Wiseman, entitled ‘Remote Sensing and Archaeology: Potential for the Future’ [11,44]. During the 1990s, RS applied to archaeology was particularly favoured by the development of several applied studies on available satellite and airborne sensors, the development of performing software and hardware and the combined use of existing technologies with the newly developed GIS (Geographical Information System). The integration of data with GIS systems led archaeologists to integrate satellite data into the concept of landscape-scale archaeology, opening up the possibility for large-scale analyses and the creation of previously unexplored spatial correlations [11,45–47]. These developments generated a real change in the RS approach applied to archaeology. This change of mindset towards a more modern interpretation by archaeologists and scholars from other disciplines to the combination of RS and archaeology has created a great development of research, materialised over time in conferences [48,49], books [11,44,50–54], reviews, and papers [46,47,55–58]. It was revolutionary for archaeology itself and for the discovery of buried cultural heritage (CH), just as it profoundly changed the way these new technologies were conceived [52,59–63].

The availability of open big-data and the development of increasingly high-performance computing and storage platforms has certainly contributed to boosting research in this field in the last few years [2,57,64–68].

Considerable progress has been made in the ability to identify a wide range of proxy indicators of the presence of buried archaeological sites. As is well known, the identification of buried archaeological features by optical satellite data is based on the observation of changes in reflectance that are useful to highlight changes: (i) in the health and phenological cycle of vegetation; (ii) and in soil moisture retention [62,69,70]. These changes are particularly evident in the red, green, near-infrared (NIR), red edge and short-wave infrared (SWIR) bands [34,71–74]. In recent years, RS studies in archaeology have focused on the use of different systems to improve the visibility of features of archaeological interest. The most common practices are (i) spectral enhancement via the creation of indices (mathematical combination between bands) such as indices derived from the use of NIR, Red, and Green (e.g., NDVI, GNDVI, and SAVI) [63,75–79] or indices based on SWIR (e.g., NDMI and MSI) [80–82]; (ii) radiometric enhancement obtained using linear and non-linear stretching or equalisation of the histogram to increase the contrast between pixel classes [83,84]; (iii) transformation, aggregation or reduction in data using various techniques such as TCT (Tasseled Cap Transformation) [85], PCA (Principal Component Analysis) and SPCA (Selective PCA) [86–89], local and global spatial autocorrelation indices (e.g., Anselin Local Moran’s I, Getis-Ord’s index and Geary’s index); and (iv) classification (e.g., K-Means, Isodata, and machine- and deep-learning based classification) [90,91].

Last but not least, a huge contribution has been provided to the development of increasingly complex machine- and deep-learning systems associated with increasingly simple and, in many cases, completely free software and tools, such as QGIS and Google Earth Engine (GEE) [1,92–100]. The entry of AI (Artificial Intelligence) into many fields, including the fields of RSA, has undoubtedly been a subject of discussion in recent years,
when systems based on pre-trained language models have entered the world scene, becoming very popular even among non-specialists, with solutions suitable for operating in various fields (e.g., graphics, text, copywriting, marketing, data analysis) [101]. These systems provide different types of output (e.g., images and text) from text or voice input by the user. For this reason, they are extremely easy to use and affordable for everyone.

The aim of this study was to analyse some aspects of the impact that an AI model based on pre-trained language models such as ChatGPT-3.5 (Generative Pre-trained Transformer) can have in the RSA, in particular the model’s ability to provide: (i) answers to (general and specific) questions on the issue; (ii) information about the references from which he/she has taken information; (iii) information about the tools to be used depending on the user’s desired outcome; and help the user to perform simple and complex processes for RSA investigations interacting with the different tools and libraries. For the purposes of the research, ChatGPT was asked to generate codes mainly in JavaScript in order to interact with the free GEE tool. In addition, the aim of the research was also to cross-sectionally assess the extent to which prior skills, competences, and language properties the user must have in order to achieve the required goal of the model.

2. Materials and Methods

The study followed the flowchart illustrated in Figure 1.

![Flowchart for the evaluation of ChatGPT-3.5.](image)

Figure 1. Flowchart for the evaluation of ChatGPT-3.5.

The following tools were used: (i) OpenAI ChatGPT-3.5 as a pre-trained language model and (ii) GEE. All conversations made between authors and ChatGPT-3.5 are shown in the SIs (Supplementary Information).
2.1. OpenAI ChatGPT-3

ChatGPT is a Generative Pre-trained Transformer (GPT) based on a natural language processing (NLP) [102–106], i.e., a large language model (LLM) that, using deep-learning, understands a text or voice input and reproduces output based on what has been understood. It was released in 2020 by OpenAI [102,107,108]. OpenAI’s main goal is to develop artificial intelligence that is safe, beneficial and accessible to all [109].

Starting in 2018, OpenAI created GPT-1, GPT-2, and released GPT-3. Version 3.5 of ChatGPT was used for this paper since it is free. It was released in 2022, and today, version 4 is already available. ChatGPT 3.5 is an upgraded version of ChatGPT 3, with several improvements in terms of accuracy, safety, and usability. ChatGPT 3.5 is generally considered to be more accurate than ChatGPT 3. This is due to a number of factors, including (i) a more sophisticated training process that uses reinforcement learning with human feedback, (ii) larger dataset of training data, (iii) improved algorithms for handling natural language. ChatGPT 3.5 is designed to be more usable than ChatGPT 3. ChatGPT-3.5 training data stopped in 2021, and many limitations are imposed on the system with regard to the language to be used in responses, the type of responses to be given, and some formal language conventions. Billions of BPE (byte-pair-encoded) tokens were used for training (Table 1).

### Table 1. ChatGPT-3.5 training data from [107].

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Tokens</th>
<th>Training Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common Crawl</td>
<td>410 billion</td>
<td>60%</td>
</tr>
<tr>
<td>WebText2</td>
<td>19 billion</td>
<td>22%</td>
</tr>
<tr>
<td>Books1</td>
<td>12 billion</td>
<td>8%</td>
</tr>
<tr>
<td>Books2</td>
<td>55 billion</td>
<td>8%</td>
</tr>
<tr>
<td>Wikipedia</td>
<td>3 billion</td>
<td>3%</td>
</tr>
</tbody>
</table>

Since ChatGPT was released, the scientific community has been using it, and articles have been published on it [107] on several topics [106,110–120], as well as ethical issues that have arisen very recently [121–126]. To date, there are few studies that demonstrate the usefulness of GPT or derived tools (e.g., Visual ChatGPT) in the field of RS and satellite image classification [127–130]. A useful tool made available by the world community via the web (e.g., GitHub or several Google extensions) is the possibility of being able to use prompts (i.e., texts explaining to ChatGPT what to do) that are already pre-compiled so as to (i) save time and (ii) prevent the system from being trained wrongly or giving wrong answers.

ChatGPT is used to obtain different types of output [106]:

(i) Generated Text: It can generate coherent and relevant responses and text based on the given questions or instructions. It can be used to answer specific questions, provide explanations, generate creative content, or even play the role of a virtual character in a conversational interaction;

(ii) Translations: By utilising ChatGPT’s translation API, it is possible to provide text in a particular language and obtain a translation in another specified language. This can be useful in supporting multilingual communication and facilitating understanding between individuals who speak different languages;

(iii) Speech Synthesis: It can be used to generate text-to-speech synthesis;

(iv) Content Research and Generation: It can be used to conduct research on specific topics and generate content based on the results;

(v) Interactive Assistance: ChatGPT’s APIs enable the creation of interactive applications that leverage its conversational capabilities to engage in conversations and respond to user queries. This can be used to develop chatbots, virtual assistants, or interactive support tools.
2.2. GEE and Sentinel-2 L2A

GEE [96] is a powerful open-source tool provided by Google. It provides a web-based interface and interactive development environment (IDE) that allows users to access and work with a wide range of datasets spanning over forty years of global data. These datasets include satellite data from missions such as MODIS, ALOS, Landsat and Sentinels, as well as other useful data such as digital terrain models, shapefiles, meteorological data and land cover information [131].

GEE is known for its high-performance computing capabilities and its ability to handle large amounts of data, making it a valuable tool in the field of remote sensing and big data analysis [94–98]. It has gained popularity in various disciplines, with the number of scientific papers on GEE increasing significantly over the years. Researchers have used GEE in several fields, such as vegetation [132–136], land use and land cover [137,138], hydrology [139–141], climate [142,143], and cultural heritage analysis [92,93,144,145].

The availability of GEE has also led to the creation and sharing of many free tools, which can be found on the GEE website [146].

For the present study, GEE was chosen because it runs in the JavaScript programming language, and its codes can be generated via ChatGPT-3.5, simultaneously overcoming the (i) problem of computing satellite data locally and (ii) writing code. The aim here is to assess the impact ChatGPT-3.5 can have on research in the field of archaeology and RS.

The data used for this study were data from the ESA Sentinel-2 (S2) L2A satellites, now considered the best in terms of spatial, spectral, and temporal resolution among the free data for RS archaeology contained in GEE [147,148]. The dataset used in all analyses is the ‘COPERNICUS/S2_SR_HARMONIZED’ dataset, which covers a time span from March 2017 to the present (Table 2).

Table 2. Sentinel-2 satellite specifications (table generated using ChatGPT-3.5, SI B, question 8).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial Resolution</td>
<td>10 m (bands 2, 3, 4, 8), 20 m (bands 5, 6, 7)</td>
</tr>
<tr>
<td></td>
<td>Coastal aerosol (Band 1), Blue (Band 2), Green (Band 3), Red (Band 4),</td>
</tr>
<tr>
<td></td>
<td>Red-edge 1 (Band 5), Red-edge 2 (Band 6), Red-edge 3 (Band 7), NIR (Band 8),</td>
</tr>
<tr>
<td></td>
<td>NIR narrow (Band 8A), Water vapour (Band 9), SWIR 1 (Band 11), SWIR 2</td>
</tr>
<tr>
<td></td>
<td>(Band 12)</td>
</tr>
<tr>
<td>Spectral Bands</td>
<td>Co. *xin r i i i. 1 (Band 1), Blue (Band 2), Green (Band 3), Red (Band 4),</td>
</tr>
<tr>
<td></td>
<td>Red-edge 1 (Band 5), Red-edge 2 (Band 6), Red-edge 3 (Band 7), NIR (Band 8),</td>
</tr>
<tr>
<td></td>
<td>NIR narrow (Band 8A), Water vapour (Band 9), SWIR 1 (Band 11), SWIR 2</td>
</tr>
<tr>
<td></td>
<td>(Band 12)</td>
</tr>
<tr>
<td>Temporal Resolution</td>
<td>5 days (global coverage)</td>
</tr>
<tr>
<td>Swath Width</td>
<td>290 km (single satellite)</td>
</tr>
<tr>
<td>Revisit Time</td>
<td>2–5 days (depending on latitude)</td>
</tr>
<tr>
<td>Data Format</td>
<td>Level-1C (Top of Atmosphere), Level-2A (Bottom of Atmosphere)</td>
</tr>
<tr>
<td>Data Source</td>
<td>Copernicus Open Access Hub ([<a href="https://scihub.copernicus.eu/">https://scihub.copernicus.eu/</a>] accessed on 25 September 2023)</td>
</tr>
<tr>
<td>Acquisition Modes</td>
<td>Global (Full Swath), Europe (Granules)</td>
</tr>
<tr>
<td>Mission Lifetime</td>
<td>7 years (planned)</td>
</tr>
<tr>
<td>Radiometric Resolution</td>
<td>12 bits</td>
</tr>
</tbody>
</table>

2.3. Approach to ChatGPT-3.5

In literature, the approach followed relies on the human-GPT interaction to generate a conversation on a given topic (e.g., RSA) and report the conversation by drawing conclusions about the validity of the model’s answers [110]. Instead, in this paper, the authors have tried to achieve a higher level of technicality and depth in order to estimate the impact of artificial intelligence (AI) in the context of RSA research. User and language model interactions were mainly based on a multi-level approach based on three stages, aimed at demonstrating the potential of the use of AI for such research for users of all levels (e.g., archaeologists with little experience in RS, archaeologists experienced in RS, engineers and programmers experienced in RS), that can be defined as follows:

(i) Entry level: (a) RS and archaeology (general theory and methods), (b) current research trends (up to 2021), (c) used references;
(ii) Medium level: (a) provide information about the tools to be used depending on the user’s desired outcome, (b) help the user to use simple functions and processes in JavaScript underlying RS applied to archaeology;

(iii) Advanced level: (a) help the user to perform complex processes for advanced RS work applied to archaeology (e.g., classification and statistics) and to create a complete process by recreating the methodology described in a scientific paper on the topic step-by-step, (c) interoperability between different tools and libraries currently used for RS in archaeology.

The evaluation also took into account the level of previous competence and the type of language used by the user to assess how the user and machine balance each other in achieving a result [149–151].

The evaluation scale used was based on a method inspired by the Likert scale, based on a psychometric technique for measuring attitude [152]. It is a multi-point accuracy scale, generally rated from 1 to 5 or 1 to 7 points, where answers are graded from the lowest value, completely wrong answer, to the highest value, completely correct answer. For the present study, a scale of 0 to 4 (5 points) was used, as follows (Figure 2):

1. Not correct (0%);
2. Almost completely incorrect. The system starts by providing minimal information, but the essential parts are missing or incorrect (25%);
3. It’s not quite correct, but it’s not totally incorrect. The system provides correct information but mistakes or omits some important information (50%);
4. Almost completely correct (75%);
5. Correct. The system explains the required topic comprehensively (100%);

**Figure 2.** Response rating scheme according to modified Likert scale.

In the case of open or descriptive answers, this scale was used in this way, where each point corresponds to a 25% increase in the goodness of the answer. In the case of questions involving a binary answer (e.g., yes and no), the values considered are 0 (incorrect) and 4 (correct).

The same rating scale was also used for the code tasks so that uniform results could be obtained for evaluation purposes. In this case, the scale was created as follows:

1. Not working (0%);
2. Partially functional or functional after intensive user interaction (25%);
3. Partially working or working after further explanation by the user at ChatGPT (50%);
4. Almost completely correct even without user interaction (75%);
5. Fully functional as generated by AI (100%);

During the process, questions or tasks with scores as shown in Figure 1 were: (i) re-submitted in a similar but not equal way; (ii) an in-depth examination or correction was
requested; (iii) changes in the answer or in the generated code were requested. This approach was repeated two to four times each time a response was unsatisfactory.

In order to ask and evaluate questions on several levels, the authors of this study have different levels of knowledge of remote sensing for archaeology. Respectively:

(i) F.V. has no experience in the use of RS for archaeology and has asked and evaluated entry-level questions. These were joined by those with more experience in order to avoid giving positive marks even in the case of wrong answers;

(ii) A.M.A. and M.S. already have experience with RS for archaeology, but none about data processing tools and therefore assessed the mid-level answers, always under the supervision of the most experienced;

(iii) N.A., M.D., R.L., and N.M. can be considered experts in RS for archaeology and assessed the answers for the advanced level.

The scores were then established by mutual agreement between the authors.

2.3.1. Entry Level

The first set of questions focused on the general use of RSA. The purpose of this approach was to assess the reliability of methodological discourse and how useful AI can be for the training of a researcher approaching the subject.

In this section, the questions posed to ChatGPT were about the theory, methods, and references of RSA, alternating with requests for more in-depth information based on the answers given by the AI itself. The topics covered are presented in full in SI A. The questions were: (i) provide an overview explaining the history of studies and the discipline from the late 19th century to 2021; (ii) deepen optical multispectral satellite remote sensing in archaeology; (iii) list the best free data for satellite remote sensing in archaeology; (iv) provide a real case study of a remote sensing study for archaeology, carried out from satellite data; (v) provide an example of a remote sensing study for archaeology, carried out with Sentinel-2 from satellite data; (vi) give information on the use of Sentinel-2 for the discovery of buried archaeological features by providing a step-by-step explanation, including tools and software to be used, where to start from and how to obtain at least the most commonly used vegetation indices. Finally, also add bibliographical references. An in-depth analysis of the relationship between AI and references was therefore analysed separately and is not part of the overall statistics. The question asked in this case was, "Please report 10 important scientific references for each year starting written in the N about ‘Remote Sensing’ and ‘Archaeology’ using scheme author(s), year, title, journal", where N is a year between 2010 and 2020 (SI D).

2.3.2. Medium Level

The second level (SI B) of in-depth study was based on the assumption that the user is already familiar with the main issues concerning RSA. The questions were, therefore, of a theoretical-practical nature and were aimed at having the user create code strings for relatively simple operations. These operations mainly concern the use of satellite data to create outputs useful for RS studies for archaeology, such as (i) RGB images, (ii) infrared false colour images, and (iii) vegetation indices. GEE was used as a tool for satellite data analysis.

The questions were structured along the same lines as previously described, as follows: (i) describe the main tools and software used as part of RS for archaeology with Sentinel-2 data; (ii) illustrate the open source tools used as part of RS for archaeology with Sentinel-2 data; (iii) describe the most commonly used packages, libraries and modules used as part of RS for archaeology in (a) Python, (b) R and Rstudio and (c) JavaScript; (iv) show the specifications of the Sentinel-2 satellite; (v) create a code base in JavaScript, compatible with GEE, to select, filter and crop the S2-L2A dataset on a geometry called an Area of Interest (AOI); (vi) display (map.AddLayer function) RGB and Infrared False Colour (R: Nir, G: Red, B: Green) annual averages on a map; (vii) finally, add to the collection the most commonly used vegetation indices for archaeological RS for a flat agricultural landscape.
2.3.3. Advanced Level

The advanced level (SI C) was a much more technical and practical approach than the previous ones. At this stage, ChatGPT was asked to reproduce a methodological approach used in other studies of RSA. Reference was made to the papers [144]. In particular, the methodology used in [92,93,144] and generally applicable to other contexts was followed.

The methodological approach used in these papers involves the following steps with the aim of improving the visibility of features of archaeological interest far beyond the possibilities offered by individual vegetation indices, i.e.,: (i) choice of dataset; (ii) choice of period of interest; (iii) dataset filtering (cloudiness and AOI); (iv) creation of vegetation indices throughout the collection considered; (v) selection, on the basis of already known or identifiable evidence of areas of archaeological interest and areas of no archaeological interest, (vi) analysis of spectra and creation of M statistic to evaluate the images in the collection where there is the greatest difference in signal, where M is described in [153]; (vii) selection of images with M > n, where n is a value close to 1; (viii) merging all images into one multi-band image; (ix) data normalisation (as suggested by ChatGPT-3.5); (x) Create a Selective Principal Component Analysis (SPCA); (xi) Calculate statistics of image neighbourhoods; (xiii) and produce a classification (unsupervised and supervised). Produce an unsupervised and supervised classification. In the latter case, ChatGPT-3.5 chose to use a K-means as the unsupervised classification [154] and an SVF (Support Vector Machine) [155,156] as the supervised one. All operations were carried out with the aim of having ChatGPT create a single JavaScript code that could be used in GEE. The aim was to prove its usefulness in the creation of complex flowcharts.

3. Results

The final results of ChatGPT-3.5’s answers to the questions asked show interesting behaviour and are shown in Table 3. The complete transcripts are shown in SIs A, B, and C.

<table>
<thead>
<tr>
<th>Level</th>
<th>Number of Questions Asked</th>
<th>Mean Score</th>
<th>StdDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entry</td>
<td>19</td>
<td>2.263157895</td>
<td>1.617083315</td>
</tr>
<tr>
<td>Medium</td>
<td>23</td>
<td>3.260869565</td>
<td>1.2420881273</td>
</tr>
<tr>
<td>Advanced</td>
<td>54</td>
<td>2.611111111</td>
<td>1.7131142</td>
</tr>
</tbody>
</table>

The result of the entry-level answers achieved a score of 43 out of a maximum of 76 and are described in Section 2.3.1. and SI A. The system was able to answer the questions posed, although it made some errors.

On the general questions about the theory (SI A, 1–6), ChatGPT-3.5 provides acceptable answers, especially for students, researchers, and scholars who want to approach the subject. It is capable of generating a credible, structured text that could easily be used as a basis for developing further research. In 7 to 19 (SI A), ChatGPT-3.5: (i) correctly cited a study done by S. H. Parcak [11], although providing a slightly wrong year of publication; (ii) incorrectly quoted works by Drs. R. Lasaponara and N. Masini, giving plausible but not true titles, although very close to the originals; (iii) gave his own interpretation of RS works for archaeology, simulating possibly real cases.

In particular, references turned out to be a problem already encountered in other papers, as ChatGPT-3.5 often provides information that is similar to the real, plausible, but not true [157–159]. The conversation is shown in SI D and can be resumed as follows (Table 4).
Table 4. Evaluation of references provided using ChatGPT-3.5, year by year, with keywords “Remote Sensing” and “Archaeology”. For each year, 10 texts were asked to be cited.

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of Completely Correct References</th>
<th>Number of Fake References</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>2011</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>2012</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>2013</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>2014</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>2015</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>2016</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>2017</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>2018</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>2019</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>2020</td>
<td>NaN</td>
<td>NaN</td>
</tr>
</tbody>
</table>

The results of an analysed sample of 100 texts produced by AI show a percentage of 99% of invented texts or texts similar to real ones but not correct. Only in one case did the AI correctly quote a text. SI D shows how GPT reworks authors, titles, years and magazines in a way that creates plausible, but not real, references [157]. For the year 2020, GPT provides no information but invites the applicant to consult Google Scholar and other scientific reference repositories.

The result for the medium-level answers was 75 out of 92, with an average of 3.26. Details are given in Section 2.3.2. and SI B.

ChatGPT-3.5 proved capable of: (i) provide general information on software used in RS for archaeology (SI B, 1–3); (ii) indicating the required code in a console to be copy-and-pasted directly into R, RStudio, Python, and GEE interfaces (SI B, 4–7); (iii) create tables from scratch with the required data (e.g., Table 2 or SI B, 8); (iv) developing simple codes such as those related to dataset selection or the selection of masks and areas of interest (SI B, 9–12) (Figure 3); (v) creating functions to generate vegetation indices (SI B, 13–23); and (iv) having the produced data displayed in the GEE map screen, such as true colour visualisation (Figure 4b), false colour infrared (Figure 4c), grey scale indices (Figure 4d), and print spectral index charts (SI B, 13–23) (Figure 5). In general, few major errors were found, but overall, the system addressed all requests.

Finally, the advanced level achieved a result of 141 out of 216, with an average of 2.61. The whole conversation is reported in SI C and explained in Section 2.3.3. The system was able to generate complex codes, in some cases committing errors and forcing the operator to intervene. ChatGPT-3.5 proved to be able to respond to and generate codes, either from existing codes or based on textual user requests, and thus generate codes from scratch. There were many mistakes made by the AI, mainly related to more complex and/or unclear user requests. In most cases, the errors were either (i) addressed after two or three requests, thus changing the way the task was requested, or (ii) resetting the conversation (e.g., SI C, 38 and 39). The system has proven to: (i) be able to give general information about the RS theory for archaeology or about sites, cities, and artefacts of archaeological interest (SI C, 1–4), albeit in some cases with errors and inaccuracies (e.g., SI C, 1–2); (ii) quickly and efficiently understand, generate, and edit complex JavaScript codes at the user’s textual request, also from methodologies described in scientific papers (e.g., [144]), selecting datasets (e.g., Sentinel-2), applying masks (e.g., cloudiness), and creating functions for creating vegetation indices and charts (SI C, 5–19); (iii) quickly create advanced level codes to apply statistical extraction functions (e.g., mean, variance, standard deviation, and M statistic) or functions on the entire data collection considered (SI C, 20–40); (iv) successfully generate code for complex functions such as Selective Principal Component Analysis (SPCA) (Figure 6a), spatial analysis (Figure 6b), and classifications (Figure 7) that are usually not easy to write in JavaScript for a beginner or mid-level user (SI C, 41–54).
The results of an analysed sample of 100 texts produced by AI show a percentage of 99% of invented texts or texts similar to real ones but not correct. Only in one case did the AI correctly quote a text.

SI D shows how GPT reworks author s, titles, years and magazines in a way that creates plausible, but not real, references [157]. For the year 2020, GPT provides no information but invites the applicant to consult Google Scholar and other scientific reference repositories.

The result for the medium-level answers was 75 out of 92, with an average of 3.26. Details are given in Section 2.3.2. and SI B.

ChatGPT-3.5 proved capable of: (i) provide general information on software used in RS for archaeology (SI B, 1–3); (ii) indicating the required code in a console to be copy-and-pasted directly into R, RStu dio, Python, and GEE interfaces (SI B, 4–7); (iii) create tables from scratch with the required data (e.g., Table 2 or SI B, 8); (iv) developing simple codes such as those related to dataset selection or the selection of masks and areas of interest (SI B, 9–12) (Figure 3); (iv) creating functions to generate vegetation indices (SI B, 13–23); and (v) having the produced data displayed in the GEE map screen, such as true colour visualisation (Figure 4b), false colour infrared (Figure 4c), grey scale indices (Figure 4d), and print spectral index charts (SI B, 13–23) (Figure 5). In general, few major errors were found, but overall, the system addressed all requests.

**Figure 3.** Example of code generated using ChatGPT-3.5. The system provides a console from which the code text can be copied. In this case, it is a JavaScript code that can be used in GEE to select the Sentinel-2 dataset at a precise time interval with a cloudiness threshold of 10%.

**Figure 4.** (a) Location of the area of interest shown in (b–d) related to the identification of sections of the Via Appia Antica, as discussed in Lasaponara et al. 2021 [144], used as a comparison study.
for the present work; (b) true colours visualisation (R: Red, G: Green, B: Blue) annual (2017–2023) average; (c) false colours visualisation (R: Nir, G: Red, B: Green) annual (2017–2023) average; (d) NDVI annual (2017–2023) average. All the images were obtained in GEE from the code generated using ChatGPT-3.5.

Figure 5. (a) Area of interest in true colour visualisation with positioning of points of interest relating to areas where features of archaeological interest have been identified and areas where there is presumably no archaeological significance; (b) graph containing the average trend over time of the NDVI index at the points indicated in a. All data were produced using GEE and the codes provided by ChatGPT-3.5; the points were selected by hand by the user, based on [144]. This data was used as the basis for the statistical calculations of the advanced level.
The system has proven to:

(i) be able to give general information about the RS theory for archaeology or about sites, cities, and artefacts of archaeological interest (e.g., SI C, 1–4), albeit in some cases with errors and inaccuracies (e.g., SI C, 1–2);

(ii) quickly and efficiently understand, generate, and edit complex JavaScript codes at the user's textual request, also from methodologies described in scientific papers (e.g., [144]), selecting datasets (e.g., Sentinel-2), applying masks (e.g., cloudiness), and creating functions for creating vegetation indices and charts (e.g., SI C, 5–19);

(iii) quickly create advanced level codes to apply statistical extraction functions (e.g., mean, variance, standard deviation, and M statistic) or functions on the entire data collection considered (e.g., SI C, 20–40);

(iv) successfully generate code for complex functions such as Selective Principal Component Analysis (SPCA) (Figure 6a), spatial analysis (Figure 6b), and classifications (Figure 7) that are usually not easy to write in JavaScript for a beginner or mid-level user (e.g., SI C, 41–54).

Figure 6. Comparison between data obtained from the code produced using ChatGPT-3.5 (a,b) and those published in [144] (c,d): (a,c) Selective PCA; (b,d) Spatial statistics on SPCA.

Figure 7. Results of the code generated using ChatGPT-3.5: (a) unsupervised classification (K-means) applied to the SPCA; (b) supervised classification (Support Vector Machine) applied to the SPCA.

4. Discussion

The conversations with ChatGPT revealed how this system has advantages and limitations for a practical application such as RS for archaeology. The numerical results (Table 3) obtained from the similar Likert scale, according to the flowchart and the evaluations expressed in Figures 2 and 3, show that there were appreciable differences depending on the use and type of questions asked.

The entry-level was the one with the lowest average total score. This was mainly dictated by the incorrect information provided in the requests for scientific papers and references. In the context of textual generation, in fact, these arguments were over-interpreted by the AI to the extent that they were credible but not usable for scientific or research work. This may be mainly due to the sources used for AI training, which probably do not include precise references on the topic of Remote Sensing Archaeology. In this case,
GPT re-processes on the basis of knowledge that it has consistent but not real elements. It is conceivable that the IA has read contributions on the subject (e.g., from Wikipedia pages or Google Books), as well as popular pages on the subject, in which there are no clear references and then reworked the topic. In fact, ChatGPT-3.5 is able to bring back the names of the major authors on the subject of RSA, as well as the main journals and the main topics addressed in the literature, albeit with its own reinterpretation.

The next two levels, i.e., medium and advanced, also showed large differences in scores. The highest score was achieved in the case of the medium level. This involved relatively simple code-writing requests, often aimed at obtaining a single result. ChatGPT-3.5 proved extremely useful, particularly in writing code related to (i) the selection and filtering of datasets, (ii) the calculation of vegetation indices, and (iii) the visualisation and export of data. The AI has proven to be a very versatile tool for writing RS-related code in different languages and for converting code from one language to another. In particular, the creation of codes for the creation of visualisation indices proved to be extremely useful as it made it possible to create several different vegetation indices, already set with the bands of the Sentinel-2 satellite, without any particular expenditure of user’s time in operations such as searching for the bands of the satellite in use for the chosen indices and writing the correct mathematical formulae in JavaScript. These operations, although not particularly complex, are often time-consuming. Another strength of RSA is that at the end of each piece of code, it provides an explanation and hints related to that code. An example is given in SI C questions 48 and 49. In these is the request to write a code to create an SPCA function. ChatGPT-3.5 suggests that the user also apply a data normalisation function in order to make the PCA itself work: “[...] Keep in mind that PCA is sensitive to the scaling of input data. Normalising the data before performing PCA, as you did earlier, is a good practice to ensure meaningful results”. Furthermore, the results show that by exposing the methodology in a textual manner to the GPT system and requiring it to generate step-by-step code, it was possible to achieve results similar to those published using other tools [144].

A marked drop in performance was observed when writing complex codes with functions linked together throughout the chat. This is mainly due to errors in the writing of the code (e.g., Python functions that cannot be used in the GEE portal) or to user requests that are not always clear or not always understood by GPT itself. Another problem encountered (e.g., SI C, 38–39) was that of recursive error. That is, within the same conversation, once a misunderstanding or an error has occurred, this is carried over throughout the series of answers. This phenomenon is clearly emblematic of the essence of ChatGPT-3.5, which, in addition to being pre-trained, learns and works from the conversation with the user. When this point was reached, it was necessary to start a new conversation and provide GPT with the code produced up to that point so that new questions could be asked. ChatGPT-3.5 was able to read, understand, and explain the provided code to the user in text language. This aspect has also proved useful in cases where the user already possesses a starting code (e.g., tutorials made available on the internet) and wants to analyse and understand its features despite having limited knowledge of the programming language.

5. Conclusions

ChatGPT-3.5 proves to be a valid tool for beginners as well as intermediate or advanced users. It is able to provide useful information about software, tools, and techniques to be used in working with RS archaeology. As shown in the previous paragraph and SIs, however, the user must be cautious in using the information as it is provided, as the data may be distorted by arbitrary interpretation generated by the AI due to the information databases it contains. In fact, the system always tries to provide an answer to the question asked, although it does not always return valid and scientifically reliable information. In particular, the biggest problems occurred when GPT was asked to illustrate a topic related to a scientific paper. In these cases, the system created a plausible argument, probably similar to the truth, but reinterpreting it. However, ChatGPT-3.5 itself advises the user to pay attention since, as GPT itself admits, it is a speech with an artificial intelligence system.
It is possible that the limitations encountered are due to privacy restrictions or an internal policy related to the laws of the different states where the system is used. As pointed out by other authors, ChatGPT-3.5 tends to give a high percentage of references with similar detail to the original (e.g., confabulated) [157].

Finally, ChatGPT-3.5 proved to be a very valuable tool in obtaining a fairly summarised overview of certain topics, such as the general themes of RS and archaeological RS, the sites of interest to be analysed and the tools that can be used. Similarly, it proved to be a useful and fast tool for generating tables on certain topics, such as Table 2, which shows data on the ESA Sentinel-2 satellites.

ChatGPT-3.5 proved to be able to respond to and generate simple and complex codes, either from existing codes or based on textual user requests, and thus generate codes from scratch. It can be a valuable support for both beginners and advanced users. In particular, it proved useful mainly for operations of a medium level of difficulty. In this segment, ChatGPT-3.5 is at its best, managing to address requests in an optimal manner and saving the user effort and time. The case of an advanced or expert user of both the programming language and RS archaeology is different. These users may find utility mainly for two operations: (i) converting code from one language to another and (ii) calling up particular functions that are difficult to write or remember (e.g., remembering bands in calculating vegetation indices). On the other hand, in the case of writing code from scratch, advanced users may probably encounter problems in using ChatGPT-3.5 and end up slowing down their work. In addition, ChatGPT-3.5 has shown that it can be used as an interpretation tool for an already-described methodology. This feature, a sort of methodological reverse engineering, could be particularly useful in the field of archaeological RS as it can enable scholars to explain a methodology described in a paper to the AI system and generate the code to replicate it, as demonstrated with the replication of the methodology of one of the reference papers [144].

ChatGPT version 3.5 is not the best performing in terms of text comprehension and response, and better performance in RS for archaeology can be achieved using ChatGPT-4, which is not free. A further increase in performance could be achieved using GPT-4 with its API (Application Programming Interface) connected to other services or using similar systems such as the connection between ChatGPT-4 and Bing or Google’s Bard. Recently, several tools and plug-ins have been developed for geoscience, including RS, based precisely on the use of the ChatGPT-4 API. An example of this is the QGIS plug-in called QChatGPT, which allows the GIS environment to be connected to AI. Certainly, given recent fast development trends, AI of this type can be implemented in the automatic identification processes of features of archaeological interest and in the classification, segmentation, and recognition of features of archaeological interest in remote sensing images. In addition, to facilitate such use, ready-to-use prompts for the RS archaeology can be created and made available to users in the same way as they currently already exist in the form of template prompts relating to a wide variety of topics (e.g., communication, social media, automated responses to emails).

In conclusion, the paper demonstrates how this tool can be carefully incorporated into the workflows of the RSA, especially for low- and mid-level users. In particular, ChatGPT-3.5, and also GTP-4, can be successfully used (i) to obtain an overview of certain issues, (ii) generate lines of code, (iii) convert codes from different programming languages, and (iv) understand already written codes in order to rework or modify them. These are all activities that can be inscribed in RS workflows for archaeology by students and researchers. Although GPT has proven to be useful, there is a need for some important considerations as a warning for users to be cautious. It is important to emphasise that, especially for entry (low) and medium levels, ChatGPT can also be a harmful tool. In fact, it must be emphasised that many of the theoretical or reference answers were wrongly given by the system, even though they were proposed to the user as true or correct. It is only the side-by-side evaluation between the non-expert user and the expert user that made it possible to understand the problem in GPT’s answer. This problem may depend on two
interrelated factors that are detrimental to the user in the case of incorrect information: (i) GPT’s propensity to always answer and very rarely admit to having no answer, and (ii) AI training datasets that may contain vague or incorrect information.

**Supplementary Materials:** The following supporting information can be downloaded at: DOI: 10.5281/zenodo.10213589; link: https://zenodo.org/records/10213589, accessed on 8 December 2023.


**Funding:** This research received a grant from the Project Basilicata Heritage Smart Lab—PO ERDF Basilicata (2014-20)—IST19013290.

**Data Availability Statement:** The data presented in this study are available in this paper and in supplementary materials (DOI: 10.5281/zenodo.10213589; link: https://zenodo.org/records/10213589, accessed on 8 December 2023).

**Conflicts of Interest:** The authors declare no conflict of interest.

**Declaration of Generative AI in Scientific Writing:** During the preparation of this work, the author(s) used ChatGPT-3.5 in order to test its operation in the field of Remote Sensing for archaeology. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

**References**

1. Agapiou, A. Remote sensing heritage in a petabyte-scale: Satellite data and heritage Earth Engine© applications. *Int. J. Digit. Earth* 2017, 10, 85-102. [CrossRef]
22. Stein, A. Surveys on the Roman Frontier in Iraq and Trans-Jordan. Geogr. J. 1940, 95, 428. [CrossRef]
23. Eaton-Thompson, G. Zimbabwe based on the British Association report. Antiquity 1929, 3, 424–433. [CrossRef]
53. Tapete, D. Remote Sensing and Geosciences for Archaeology. Geosciences 2018, 8, 41. [CrossRef]
68. Kitchin, R. Big Data, new epistemologies and paradigm shifts. *Big Data Soc.* **2014**, *1*, 1. [CrossRef]


159. Day, T. A Preliminary Investigation of Fake Peer-Reviewed Citations and References Generated by ChatGPT. *Prof. Geogr.* 2023, 1–4. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.