

Correlating the Allies-of-Humanity message with the UFO research literature

May 2023

Abstract

The goal of this work was to match, using NLP tools, the content of the Allies-of-Humanity (AoH) message, transmitted to the world through Marshall Vian Summers, vs the contents of a selected set of UFO research books. The task was to identify the books, chapters, and paragraphs, which are most close to the AoH message content as a whole or to its parts. The NLP tools used were standard methods for measuring similarity of textual documents: cosine similarity of documents represented as bag-of-words or TF-IDF vectors, and soft cosine similarity of the document vectors, taking into account the between-terms-similarity according to word2vec or fasttext model. To overcome the tendency of paragraph matching to select pairs of short one-sentence paragraphs with literally identical words, a simple algorithm was used to preliminary enlarge the paragraphs via merging similar adjoining ones.

With varying degree of contrast, usage of different documents similarity models highlighted within the UFO research literature (specifically, concluding sections) similar themes, most relevant to AoH message: the global implications of the on-going contact with the ET presence on Earth for the evolution and the future of humanity; the nature of alien intentions; spiritual/metaphysical connection. Notable is high relevance of the UFO abductions literature.

Project code : <https://github.com/hdnws/aoh-ufo>.

1 Introduction

1.1 Motivation

While technically this article pertains to the field of the natural language processing, the nature of text material chosen for study necessitates a somewhat lengthy overview of the UFO research field.

There exists widespread attitude to the UFO/UAP phenomenon, depicting the whole subject as non-existing and illusionary, attributing the reported observations and cases of contact to misperception of known phenomena, products of imagination or deliberate hoaxes. That attitude tends to subsume if one

takes the effort of looking at the gathered documentary evidence – the many thousands of military and intelligence reports, composed by qualified observers, psychologically stable persons, where some part of the cases is confirmed with the instrumental means of objective control, such as photographs, videotapes and radar data [25, 24]. Citing the famous statement by the Roscoe Hillenkoetter, the first CIA director: “Behind the scenes, high-ranking [US] Air Force officers are soberly concerned about the UFOs. But through official secrecy and ridicule, many citizens are led to believe the unknown flying objects are nonsense”. [21, p. 289]. In the USA, the country with a remarkable level of observed UFO activity, this official policy of debunking and ridicule of UFO was systematically implemented from the early post-WWII years with their waves of mass sightings, and from then on is continuously reinforced by the press and media. [24, Appendix 5]. A “scientific” sanction for such policy was produced by the CIA-sponsored Robertson panel of Nobel-caliber scientists in 1953 and USAF-sponsored Condon committee report in 1968 [26, chapters 4, 8]. Similar policy was adopted in many other countries, interestingly including those not under USA influence – as in USSR [21, chapter 10].

Leaving aside the deliberately distorted picture of the UFO phenomenon, produced by the controlled [56, 57] media, it is worthwhile to cite several available official reports:

- “The phenomenon reported is something real and not visionary or fictitious” – N. Twining memorandum, 1947 [23, p 113]
- “These phenomena are obviously the result of a high technical skill which cannot be credited to any presently known culture on Earth” – USAF Europe report, 1948. [23, p 115]
- “a. The matter is the most highly classified subject in the United States Government, rating higher even than H-bomb. b. Flying saucers exist. c. Their modus operandi is unknown but concentrated effort in being made by a small group headed by Doctor Vannevar Bush. d. The entire matter is considered by the United States authorities to be of tremendous significance.” – Top Secret 1950 memorandum from Wilbert Smith, Senior Radio Engineer with the Canadian Government’s Department of Transport. The information was obtained from Dr Robert I. Sarbacher, an American consultant to the Joint Research and Development Board [23, p 133]:
- “Unidentified flying objects—sometimes treated lightly by the press and referred to as ‘flying saucers’—must be rapidly and accurately identified as serious USAF business in the ZI [Interior Zone]. ... Technical and defense considerations will continue to exist in this area.” – the 1959 memorandum of the USAF inspector general to the air base commanders in the continental US [21, p 288].

The unidentified objects exhibit unattained to present day capabilities: hypersonic flight speeds exceeding thousands mph, rapid right-angle turns at high

speed, nearly-instantaneous change of elevation, switching on and off invisibility for visual and radar observation, the free cross between air and water environment. The objects made numerous intrusions to sensitive sites and no-fly-zones, in particular related to nuclear arms and strategic defense installations, sometimes causing equipment malfunction [23, chapter 18].

Bottom line: humanity faces the presence of unknown – alien – intelligence(s), of which technological capabilities vastly outperform our own.

The implications are profound. Citing the declassified 1968 report “UFO Hypothesis and Survival Questions” by an undisclosed NSA analyst: “If ‘they’ discover you, it is an old but hardly invalid rule of thumb, ‘they’ are your technological superiors. ... Human history has shown us time and again the tragic results of a confrontation between a technologically superior civilization and a technologically inferior people.” A weaker culture in such a contact faces the very real dangers of the loss of identity and absorption by the others. [21, p 423], [26, chapter 8].

Whether such a threat fulfills depends on the ultimate goals of the visitors. While obviously different visitors’ groups could have different motives, some keys to understanding the aliens’ intentions can be obtained via observing their actions. Available UFO research literature allows to list the following typical activities of UFOs and their occupants:

- Monitoring of the Earth’s biological environment - taking soil/plants samples [53, chapter 1].
- Surveillance of the human activity, in particular - military/defense installations, nuclear storage/production sites, space probes [28]; this occasionally involves interference into defense/space missions.
- Imparting in person messages to select human individuals - ‘contactees’ - for dissemination to human society at large. While there exists indirect evidence, confirming that such contacts did have place, and psychological evaluation of some of the witnesses sees them as honestly reporting their experiences, part of the messages turns out to be disinformation [21, chapter 14. CIA, section ‘disinformation’]. A variant of imparting information to a select individual via mental communication without physical presence is known as ‘channeling’, same considerations of disinformation applied.
- Imparting impersonal messages to human society via crop circle drawings [47, 48]
- Harvesting of biological resources of Earth – known as ‘cattle mutilations’ – removing the blood and internal organs of (mostly) domestic cattle and wild animals [42]. There exist isolated undocumented reports of mutilated humans [41, chapter 14].
- Taking people against their will – known as ‘abductions’; taking their sperm and ova for the purpose of creating hybrid beings; suppressing the

memories of the subjects; putting tiny implants into them, presumably for tracking and monitoring [32]–[39]

The last two aspects of UFO activity are the most disturbing. It must be stressed that both the alien being themselves and the hybrid beings which they create with the program of abductions, possess superb mental capacities, by large exceeding the known Psi capacities of select humans: they can read peoples' minds, can extract past memories, they can suppress memories, causing a subject not to want remembering his/her experience; they can supplant false impressions/memories/ideas into one's mind; finally, and most important, they can control humans mentally, even without the subject realizing him/her being controlled.

These aspects of manipulated subjects' memories has put the abduction research under severe criticism, especially given that part of memories of abduction experiences could be retrieved only under hypnosis. But, first, physical evidence of abduction phenomenon confirms its reality: people are temporarily missing, they get inexplicable scars and scoop marks, sometimes the process of abduction is witnessed by external observers; a number of implants have been surgically removed and undergone scientific study, which, among other, found isotopic ratios of compound metals different from those on Earth [41]. Second, the testimony of multiple abductees is consistent vs each other; that includes subjects who were totally unaware of UFO field (such as small children), ruling out the explanation of abduction cases as simple retelling of stories from the popular culture.

David Jacobs, American historian, who has researched the abduction phenomenon probably more than anyone else, has come to startling conclusions about the ultimate intentions of the aliens, perpetuating it:

All the evidence seems to suggest that integration into human society is the aliens' ultimate goal. And all their efforts and activities appear to be geared toward complete control of the humans on Earth. Indeed, the abductees are already living with the burden of alien visitation and manipulation.

It is now possible to discern at least four specific programs that the aliens have put into effect to achieve their goal:

1. The Abduction Program. The aliens initially selected human victims around the world and instituted procedures to take these humans and their progeny from their environments without detection.
2. The Breeding Program. The aliens collect human sperm and eggs, genetically alter the fertilized embryo, incubate fetuses in human hosts, and make humans mentally and physically interact with the offspring for proper hybrid development.
3. The Hybridization Program. The aliens refine the hybrids by continual alteration and breeding with humans over the generations to become more human while retaining crucial alien

characteristics. Perhaps humans are also altered over time and acquire alien characteristics.

4. The Integration Program. The aliens prepare the abductees for future events. Eventually, the hybrids or the aliens themselves integrate into human society and assume control. [35, chapter 12. Life As We Know It?]

Planetary Acquisition

...Abductees describe a program of well-ordered, nonviolent (with individual abductee exceptions), gradual, stealthy takeover. When this integration program succeeds, it will mean the end of humans' control over their own destiny. If successful, it strongly suggests, as I wrote in *The Threat*, the end of *Homo sapiens* and the beginning of a new species of human-*Homo alienus* (hybrids and the offspring of hybrids and humans). Physically, these creatures will be exactly like *Homo sapiens*; but they will be endowed with different neurological capabilities. [36, chapter 10. Integration and Speculation]

The texts of selected UFO research literature make the first sub-corpus of this NLP study.

The second sub-corpus is the 'channeled' information - the Allies-of-Humanity message passed through Marshall Vian Summers in 1997-2016. With all due caution which pertain to this kind of information, it is remarkable how AoH message matches the scenario of a stealth planetary takeover, which David Jacobs has reconstructed by 'reverse-engineering':

The Briefings reveal that our world is undergoing an extraterrestrial Intervention by forces who, as demonstrated by their actions, are here to subvert human authority and to integrate into human societies for their own advantage. These forces represent non-military organizations who are here to seek human and biological resources. The Allies refer to these forces as the Collectives. The Collectives do not value human freedom.

Because the Intervention is being carried out by small groups of intervening forces, it must rely primarily upon deception and persuasion to achieve its goals. The Allies Briefings describe in detail how this is being accomplished and what we must do to stop it.

This extraterrestrial Intervention is being focused in four arenas:

- It is influencing certain individuals in positions of power and authority in government, commerce and religion to cooperate with the Intervention through the promise of greater wealth, power and technology.
- It is creating hidden establishments in the world from which the Intervention can exert its influence in the mental environment,

seeking to make people everywhere open and compliant to its will through a “Pacification Program”.

- It is manipulating our religious values and spiritual impulses in order to gain human allegiance to their cause.
- It is taking people against their will, and often without their awareness, to support an interbreeding program designed to create a hybrid race and a new leadership who would be bonded to the visitors. [11]

Either of these two sources, if taken seriously, by the global implications of what they describe, provides enough motivation to look closer into their content.

As for the handy label “conspiracy theory”, which is often invoked in this context, we remind that this label was introduced and reinforced into the public discourse by USA’s CIA, as documented by the declassified CIA dispatch #1035-960; initially to manipulate the public opinion regarding the Warren Commission report on JFK murder; see excellent work [58] for details.

1.2 Technical task

The goal of this NLP study was to match the content of AoH message vs the available UFO research literature. In terms of information retrieval systems, the AoH message (in form of its 12-points summary, or its parts, or as a whole) here was used as a ‘query’ to search through the corpus of UFO books/articles, locating the books/sections/fragments with the content most close to the ‘query’.

The matching fragments thus found could not be claimed as an ‘ultimate proof’ of AoH message validity: first, because of the scientific standards for ‘proof’ typically require a reproducible controlled experiment, which is not the case with UFOs (and, we note, in other fields of study of non-precise nature); second, because of certain crucial elements of UFO situation— such as Psi-phenomena/mental telepathy – are not officially acknowledged as existing (even if studied with government funding and obtaining spectacular results, [45]). The idea was rather that the found text correspondences to serve as illustration of the relevancy of the AoH-message to what has been already established independently by many UFO researchers during decades of painstaking field work.

2 Related Work

The author was unable to find existing studies of applying NLP techniques for the task of comparing AoH briefings vs UFO literature. Comparison results are available as human estimates. Many UFO researches and writers are aware of AoH message, and some have expressed their opinion in support of its validity, based on their own extensive knowledge of UFO field; [12] includes such comments from Jim Marrs and Timothy Good. Others are more cautious; e.g. David Jacobs and Richard Dolan did not mention AoH briefings, notwithstanding the fact, that these researchers investigate very close themes. Interestingly,

Jacobs in [36, chapter 10] provides detailed arguments for a specific point accentuated by AoH: namely that the Earth might be far not the first world being taken over; that this is a routine process, which has happened countless times in the universe.

3 Related methodology

There exist a large family of approaches for comparing similarity of texts.

The survey by Kilgariff [1] provides detailed overview of those of them which are based on various statistics of word frequencies: χ^2 -test, Mann-Whitney ranks test, t-test, mutual information, TFIDF. They show good results for answering question: how likely it is that the two given texts come from the same homogeneous source (corpus)?

However, this is not the question asked in this study: here the sources are essentially different. Different terms are used in the 2 sub-corpora for key concepts: in the UFO books the terms ‘UFO’ and ‘alien’ are abundant, while in the AoH-message these are rather rare, the ETs are referred primarily as ‘visitors’. So it was essential for the method of comparison of the text document to take into account the terms similarity.

Another survey by Goma and Fahmy [2] categorizes the methods of evaluating texts similarity into String-based, Corpus-based and Knowledge-based.

String-based measures operate on string sequences and character composition. They sub-categorize into character-based and term-based. The character-based methods evaluate the similarity between terms based on various kinds of distances between terms as sequences of characters. Suitable for finding similarity between morphological forms, this method again does not handle the similarity of synonyms. Among the term-based methods, the cosine-similarity and Euclidean distance, measuring the similarity of terms represented as vectors in an auxiliary semantic space, are particularly suitable for the present task.

Knowledge-Based methods find the degree of similarity between words using information derived from semantic networks [3]. WordNet is the most popular semantic network in the area of measuring the Knowledge-Based similarity between words in English. Knowledge-based similarity methods account for semantic similarity and semantic relatedness, the latter includes extra similarity relations such as is-a-kind-of, is-a-specific-example-of, is-a-part-of, is-the-opposite-of.

Corpus-based methods infer semantic similarity measure between terms according to information gained from large corpora. **Hyperspace Analogy to Language (HAL)**, and **word2vec** [4] represent terms as vectors in a multi-dimensional space based on word co-occurrences; **FastText** [5] additionally involves the sub-word information. **Pointwise Mutual Information - Information Retrieval (PMI-IR)** is also based on word co-occurrences [6]. **Latent Semantic Analysis, LSA** perform SVD decomposition of the occurrence matrix – the document-term matrix of which rows represent terms and columns – documents; the documents’ similarity is then computed as cosine similar-

ity between transformed columns [7]. **Generalized Latent Semantic Analysis (GLSA)** represents the document vectors as linear combinations (with document-term matrix as coefficients) of term-vectors; the GLSA approach can combine any kind of similarity measure on the space of terms with any suitable method of dimensionality reduction; unlike LSA, GLSA is not based on bag-of-words document vectors [8]. **Soft-cosine measure** [9, 10] calculates similarity between documents as cosine measure of their term-frequency- or tfidf-vectors, but effectively replacing the product of vector components corresponding to two terms with the similarity of those terms; where term similarity might be supplied via different methods.

In this study, soft cosine measure was chosen for comparing text-to-text similarity for the following reasons: it takes into account terms similarity; it showed state-of-the-art results in the previous work [10]; a ready efficient implementation of soft-cosine measure was available in the open-source gensim library, incorporating the term similarity from Word2Vec or FastText model.

4 Model Description

The comparison of sub-corpora texts was performed on the level of books, sections and (combined-by-similarity) paragraphs, meaning that the documents compared were books/sections/paragraphs; this introduced the 1st hyper-parameter – structure level. The documents were represented as either bag-of-words vectors, or their TFIDF transformations – this gave the 2nd hyper-parameter – frequency mode: BOW or TFIDF. The similarity between terms was calculated via training on the texts of entire dataset either Word2Vec or FastText model, adding 3rd hyper-parameter - the term similarity model. The similarity of the documents was calculated as soft-cosine measure of their bag-of-words- or TFIDF- vectors with term similarity coming from the Word2Vec or FastText model, trained on the dataset corpus. For reference, documents of each sub-corpus were also compared vs the complement sub-corpus taken as a single document (with combined complement sub-corpus as text).

4.1 Paragraph merge

When comparing the texts paragraph-to-paragraph, the pairs with top similarity turned out to be short one-sentence (or even single-word exclamation) paragraphs, composed of the same or related words, mostly from (self-)dialogues. Such results made little sense, since these small fragments did not encompass the deeper meaning of texts at all. Apparently, it was necessary either to exclude too short paragraphs from comparison, or to unite adjoining paragraphs with similar content into larger fragments before comparing them. Both options were implemented.

For uniting similar adjoining fragments a simple algorithm was devised.

1. Take the text paragraphs as the initial text fragments.

2. For each fragment calculate/update its ‘internal similarity’ as the average soft-cosine measure between sentences contained in it. For single-sentence fragments take internal similarity to be 0.
3. For pairs of adjoining fragments calculate/update:
 - the ‘between-fragments-similarity’ as the average soft-cosine measure of pairs of sentences between fragments.
 - merge score: the difference between ‘between-fragments-similarity’ and the minimal (over pair) ‘internal similarity’.
4. Find the pair with the maximum merge score. If it is positive - merge the adjoining fragments and repeat from step 2, otherwise stop.

The idea of this algorithm is that the initial composition of paragraphs of sentences provides a useful estimate – a lower bound – on how similar sentences should be to retain certain topic inside the paragraph. United are the adjoining fragments for which their similarity to each other exceeds this lower bound by largest margin.

Applying this algorithm, the number of paragraphs in the entire corpus was reduced from ≈ 56600 to ≈ 25000 for BOW frequency mode and ≈ 31000 for TFIDF frequency mode.

Matching merged paragraphs using the BOW frequency mode showed a tendency to assign top similarity to pairs of overly long paragraphs, compared with TFIDF frequency mode.

5 Dataset

The involved dataset is a collection of text documents [11]-[55], which are mostly books. The documents are structured into the sections(chapters) and sometimes subsections. The dataset consists of the two main parts:

1. The Allies-of-Humanity message. This includes the 5 upper-level items: the short 12-points summary [11] and the 4 volumes of the Allies messages (briefings), transmitted (channeled) via Marshall Vian Summers in the years 1996 to 2016 [12], [13], [14], [15].

2. UFO research books/articles. This includes 40 upper-level items, which are mostly books, with the exception of the James McDonald article. The intent was to include the works, which made important contributions into understanding of the various aspects of UFO phenomenon, especially those which gathered sufficient amount of solid data so as to enable making tentative conclusions. By no means this is an exhaustive list; many important works were not included, mostly due to lack of time for the dataset preparation. The main topics presented in this dataset texts are:

- the general overview of UFO phenomenon and related subjects [16], [17], [18];

- UFO reality as given in the official military/intelligence reports [19], [20], [21], [22], [23], [24], [25], [26], [27];
- scientific investigations of UFO [28], [29], [30], [31];
- human abductions by UFO, hybridization, implants [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [17];
- animal mutilations [42], [43];
- underground bases [54], [51];
- crop-circles [47], [48];
- connection with Psi phenomena, telepathy, spirituality [44] [45], [17], [18].

Electronic versions of most of them are present in the open electronic libraries, such as web.archive.org, scribd.com, avalonlibrary.net, etc.

The Allies-of-Humanity (AoH) message part of the dataset is further referred to as 'TARGET', the remaining part - UFO research books/articles - as 'SOURCE'. These labels reflect the task set: attempt to find the subset of the 'SOURCE' texts, which are most relevant to the 'TARGET' texts. There is no train/validation/test split of data provided, because of there could not be assigned unambiguous labels 'relevant'/'irrelevant' regarding the relation between a specific 'SOURCE' text and the entire 'TARGET' text. Though it was expected that the UFO research texts, containing general conclusions about UFO abduction phenomena (such as [35], [36]) would have high relevancy, the overall context of AoH message is clearly wider.

5.1 Preprocessing

Making the dataset involved manual and automated steps.

- Manual steps:
 - Find the electronic version of a book/article in an open electronic library.
 - Extract the book/article text from .pdf, .djvu, .epub, .mobi formats using tools such as python module pdfminer, DjvuLibre, Calibre2 ebook-convert.
 - Split the text into sections(chapters) and, sometimes, subsections via inserting lines containing a row of '=' or '-' symbols. The line after the split line was modified so as to contain the title of the section/subsection for later reference.
 - Remove the text of running titles, containing page number, book/chapter title.
- Automated steps:

- Remove sections with bibliography, references, notes, index (using section/subsection titles assigned during the manual preprocessing step). These are special kinds of text (including many named entities and numbers) with statistical properties substantially different from the main text, which was the object of analysis.
- Remove end-of-paragraphs references such as '[12]' or '12', which negatively affected NLTK tokenization to sentences - the reference number sticking to the beginning of the next sentence (this also adversely affected splitting text to paragraphs). For the same reason remove slanting double quotation marks, convert slanting single quotation marks to vertical ones.
- Dehyphenate: restore the known words, split into parts with or without hyphen over the new line; the list of known words was built from the dataset vocabulary, it included the words which occurred (in full form) minimum 2 times.
- Convert token case to the most frequent case form of a particular token (so as to reduce vocabulary size, but at the same time to preserve sentence case for named entities and thus enhance readability of the analysis output).
- Convert to 'embedded line-sentence' format to accelerate further analysis with gensim library. Here 'embedded line-sentence' means that the text of each section/subsection is converted to gensim line-sentence format: one sentence occupies one line, words separated with space, punctuation removed; additionally, paragraphs separated from each other by an empty line.

A small python framework was written specifically for processing text corpora, stored in the text files split to sections/subsections as described above: it supports iterating, with filtering, over a corpus on the different levels (book/section/subsection), and transforming of the contained texts.

5.2 Dataset statistics

The table 1 describes statistics of the dataset after preprocessing step.

	TARGET	SOURCE	TOTAL
Books/articles	5	40	45
Sections	32	848	880
Tokens	110 308	4 463 817	4 574 125
Vocabulary size	3067	41127	41192
Min word count		2	

Table 1: Statistics of the AoH-UFO-dataset. The min word count value defines minimal number of occurrences for a token to be included into vocabulary.

5.3 Dataset access

The AoH-UFO dataset is published along with this study project code at https://github.com/hdnws/aoh-ufo/corpus_raw.

6 Experiments

6.1 Baselines

The logical simple baseline to use was comparing the same documents without taking into account the terms similarity. This was done via comparing the documents, represented as the same bag-of-words- or TFIDF-vectors, using cosine similarity. This is equivalent to calculating soft-cosine measure of documents with terms' similarity being identity matrix (all terms orthogonal/independent).

6.2 Metrics

As stated earlier, the task of matching texts between two different corpora does not assume a natural metric for measuring how good a particular text-to-text match is, compared to an alternative match.

In this study the models' output was the similarity of a particular target (AoH) text (united target text, a book, a section, a [merged] paragraph) to a particular source text (UFO books/articles) – of the same structural level, or to united text.

Based on the human understanding of the text material under study, it was expected for a good model to estimate as high the similarity of the AoH message to the abduction research texts, specifically the chapters with conclusions, and to the texts concerning the global implications for humanity of the contact with ET. That expectation was met even for simple baseline models not accounting for term similarity.

For the models working on the structure level of book/section, for each target text was formed the ranked by similarity list of the source texts most close to it. On the book level this ranking included all source books, as there were few, only 40. On the section level the ranking list was limited by 64 top items. On the book/section level the model for reference put into ranking lists 2 additional statistics: 1. The similarities vs aggregated items: entire target/source text (labeled '< *TARGET* >', '< *SOURCE* >'), compared to normal/aggregated texts of the complement sub-corpus. 2. The average similarity of the given target text vs source texts on this structural level (labeled '< *SOURCE_MICRO* >'). These statistics helped to estimate sensitivity of a particular model as the difference between the similarity of the top-ranked matches with that of an average (random) match.

On the level of matching the [merged] paragraphs, the output did not include matches for all target paragraphs, since there would be too many, over 1 000; instead the top by similarity pairs of target-to-source paragraph matches were listed. Similarly, for reference was listed the average similarity of a target vs a

source paragraph – the pair ('< *TARGET_MICRO*;', ' *SOURCE_MICRO*'), and the average similarity of the aggregated target text vs source paragraphs – the pair ('< *TARGET* >', '< *SOURCE_MICRO* >').

6.2.1 Hyper-parameters

The combinations of hyper-parameters together with the baselines involved produced in fact $3 * 2 * 2 = 12$ models in this study:

- structure level $\in \{\text{book, section, paragraph}\}$
- terms similarity model $\in \{\text{none, word2vec, fasttext}\}$
- frequency mode $\in \{\text{bow (bag-of-words), tfidf}\}$

6.2.2 Model-to-model comparison

To provide a numerical estimate of how consistent results of different models were, for the book/section level models the Kendall tau rank coefficient was calculated to compare model-to-model rankings. For each pair of models compared, the average of the ranking-to-ranking comparison was calculated. The average was taken over the list of target texts, for each of which a ranking of source-texts-similarity to it was produced. The average calculated was arithmetic for the Kendall ranking correlation coefficient and geometric for the associated p-value.

For paragraph-level models such comparison was not applicable, since the list of merged paragraphs compared was model-dependent.

6.3 Experiment Setup

The texts have been split to sentences and words using methods `sent_tokenize` and `word_tokenize` from NLTK library, version 3.7. Into vocabulary were included the tokens, which occurred with minimum count 2, which had one of the following forms: 1. Sequences of English alphabet letters. 2. Sequences of digits. 3. Tokens of form 1, followed by hyphen and form 1 or 2 - to include named entities like MJ-12 or F-16. 4. Sequences of English alphabet letters with a single " " sign inside or at the end.

The most important hyper-parameters, which produced the variety of models employed, are listed in 6.2.1. The rest concern training of the Word2Vec and FastText models on the corpus texts for obtaining the terms similarity: gensim library version: 4.2.0; vector_size: 64; window: 5; min_count: 2; epochs: 10.

7 Results

The results of matching texts from source sub-corpus vs those from target sub-corpus were produced depending on the text structure level. In case of book/section level, for each book/section text from the target sub-corpus a

ranking was built with the most similar to it texts of the same level from the source sub-corpus. In case of paragraph level, listing rankings of top-similar source paragraphs for each target paragraph was impractical, so only most-similar pairs (target-paragraph, source-paragraph) were logged and reviewed.

Below are listed the top-10 rankings of the texts from the source sub-corpus vs the target sub-corpus taken as a single document for cosine similarity and for soft-cosine similarity models. They are grouped by the frequency mode used - whether the documents are represented as bag-of-words vectors, or their TFIDF-transformations.

7.1 Book level

The top similarity rankings of source book-level texts against the entire target corpus are given in the tables 2 and 3

Clearly, the TFIDF baseline model is more sensitive, the similarity of top-matches differs more significantly from that of the random match (`<SOURCE_MICRO>`), compared to bag-of-words models.

It is visible that different models ranked as top similar to target sub-corpus several repeating source texts. This is a point in support of consistency of the outputs of different models.

To measure quantitatively consistency between models, the Kendall correlation of model-to-model rankings was calculated and averaged (over the list of the target texts being matched). The results in the table 4 for the book-level texts indicate that the rankings produced by all models correlate positively; the correlation is more significant between models using the same frequency mode (bag-of-words or TFIDF).

7.2 Section level

Comparison of section-level texts provides more detailed picture of what specific content of UFO-research books/articles correlate highly with AoH message. Clearly, the most relevant sections are concluding (or, sometimes, introductory) chapters, which ask, or evaluate possible answers to, very simple but profound questions: Who the visitors are? What are their goals on Earth? What does this portend for the future of humanity?

The top similarity rankings of source section-level texts against the entire target corpus are given in the tables 5 and 6. Here the source text is identified by path: book title / section(chapter) title.

Again, same as for book-level models, for section-level models it is clear that several source texts repeatedly appear in the top rankings of similarity-to-target produced by different models, implying a degree of between-model consistency.

Same as for book-level models, for section-level models the average Kendall-tau correlation coefficient between models rankings was calculated, as shown in the table 7 .

Title	Similarity
Model - Cosine Similarity, term similarity - none	
Dolan R. UFOs for the 21st Century Mind, 2014	0.869
Jacobs D. The Threat, 1998	0.864
Swann I. Penetration, 1998	0.864
Hopkins B, Rainey C. Sight Unseen, 2003	0.859
Jacobs D. Secret Life, 1993	0.856
Jacobs D. Walking Among Us, 2015	0.851
Andrews C. Crop Circles-Signs of Contact, 2003	0.847
Mack J. Abduction-Human Encounters with Aliens, 1994	0.846
Sauder R. Underground Bases And Tunnels, 2014	0.845
<SOURCE>	0.843
Friedman S. Flying Saucers and Science, 2008	0.842
<SOURCE_MICRO>	0.820
Model - Soft Cosine Similarity, term similarity - word2vec	
Swann I. Penetration, 1998	0.909
Hopkins B, Rainey C. Sight Unseen, 2003	0.897
Dolan R. UFOs for the 21st Century Mind, 2014	0.897
Jacobs D. The Threat, 1998	0.894
Jacobs D. Secret Life, 1993	0.885
Friedman S. Flying Saucers and Science, 2008	0.885
Jacobs D. Walking Among Us, 2015	0.877
<SOURCE>	0.875
Mack J. Abduction-Human Encounters with Aliens, 1994	0.875
Howe L. An Alien Harvest, 2014	0.875
Hynek A. The UFO Experience-A Scientific Inquiry, 1972	0.874
<SOURCE_MICRO>	0.846
Model - Soft Cosine Similarity, term similarity - fasttext	
Swann I. Penetration, 1998	0.888
Dolan R. UFOs for the 21st Century Mind, 2014	0.880
Hopkins B, Rainey C. Sight Unseen, 2003	0.880
Jacobs D. The Threat, 1998	0.871
Friedman S. Flying Saucers and Science, 2008	0.870
Jacobs D. Secret Life, 1993	0.866
Howe L. An Alien Harvest, 2014	0.860
<SOURCE>	0.859
Hynek A. The UFO Experience- A Scientific Inquiry, 1972	0.858
Corso P. The Day After Roswell, 1997	0.857
Good T. Alien Contact, 1991	0.856
<SOURCE_MICRO>	0.828

Table 2: The rankings of similarity of SOURCE **book**-level texts to the aggregated TARGET text; documents – **bag-of-words** vectors, term similarity – none, word2vec, fasttext.

Title	Similarity
Model - Cosine Similarity, term similarity - none	
Dolan R. UFOs for the 21st Century Mind, 2014	0.0741
<SOURCE>	0.0685
Marrs J. Alien Agenda, 1997	0.0498
Good T. Earth-the aliens enterprise, 2013	0.0412
COMETA report. UFOs and Defense-What Should We Prepare For, 1999	0.0405
Mack J. Abduction-Human Encounters with Aliens, 1994	0.0388
Dolan R. UFOs and the National Security State-V2, 2009	0.0375
Hynek A. The UFO Experience-A Scientific Inquiry, 1972	0.0368
Jacobs D. Secret Life, 1993	0.0353
Corso P. The Day After Roswell, 1997	0.035
Good T. Above Top Secret, 1988	0.0348
<SOURCE_MICRO>	0.0258
Model - Soft Cosine Similarity, term similarity - word2vec	
Dolan R. UFOs for the 21st Century Mind, 2014	0.369
Jacobs D. The Threat, 1998	0.267
Jacobs D. Walking Among Us, 2015	0.257
<SOURCE>	0.256
Marrs J. Alien Agenda, 1997	0.253
COMETA report. UFOs and Defense-What Should We Prepare For, 1999	0.252
Hynek A. The UFO Experience- A Scientific Inquiry, 1972	0.239
Mack J. Abduction-Human Encounters with Aliens, 1994	0.226
Swann I. Penetration, 1998	0.225
Jacobs D. Secret Life, 1993	0.221
Corso P. The Day After Roswell, 1997	0.219
<SOURCE_MICRO>	0.160
Model - Soft Cosine Similarity, term similarity - fasttext	
Dolan R. UFOs for the 21st Century Mind,2014	0.421
Hynek A. The UFO Experience- A Scientific Inquiry, 1972	0.364
<SOURCE>	0.353
COMETA report. UFOs and Defense-What Should We Prepare For, 1999	0.353
Marrs J. Alien Agenda, 1997	0.339
Friedman S. Flying Saucers and Science, 2008	0.306
Jacobs D. The UFO Controversy in America, 1976	0.300
Corso P. The Day After Roswell, 1997	0.297
Jacobs D. The Threat, 1998	0.293
Jacobs D. Secret Life, 1993	0.291
Mack J. Abduction-Human Encounters with Aliens, 1994	0.287
<SOURCE_MICRO>	0.237

Table 3: The rankings of similarity of SOURCE **book**-level texts to the aggregated TARGET text. Documents – **TFIDF** vectors, term similarity – none, word2vec, fasttext.

Model 1/2	(w2v,bow)	(ft,bow)	(none,tfidf)	(w2v,tfidf)	(ft,tfidf)
(none,bow)	0.63, 2.5e-9	0.66, 5e-10	0.25, 1.9e-2	0.46, 1.4e-05	0.38, 4.0e-4
(w2v,bow)		0.89, 9e-17	0.30, 5.3e-3	0.56, 1.5e-7	0.41, 1.0e-4
(ft,bow)			0.30, 4.0e-3	0.53, 5.2e-7	0.43, 4.2e-5
(none,tfidf)				0.61, 1.2e-8	0.72, 2.0e-11
(w2v,tfidf)					0.69, 1.1e-10

Table 4: The averaged Kendall-tau correlation between rankings of book-level models. Each row/column represents a similarity model, the cell of row/column intersection – results of comparison of rankings of 2 models. In each cell the 1st number is the arithmetic average of Kendall correlation coefficient, the 2nd number is the geometric average of associated p-value. Models’ abbreviations: none – cosine similarity baseline model; w2v/ft – soft-cosine similarity model using word2vec/fasttext terms similarity. bow/tfidf - frequency mode used for the document vectors.

7.3 Paragraph level

Matching of paragraphs was performed using only the soft-cosine similarity models. The paragraph documents again were represented either as bag-of-words vectors or their TFIDF transformations; the term similarity was found from either word2vec or fasttext model trained on the entire corpus. Before comparison, as was described in 4.1, the paragraphs were enlarged via merging together the adjoining similar ones, using a particular soft-cosine similarity model. This caused each soft-similarity model to compare specific to it lists of paragraphs. Experiments showed that the models representing the paragraphs as bag-of-words vectors assign top similarity to overly large paragraphs, attributing concise meaning to which was difficult. The 2 models, representing the paragraphs as TFIDF vectors, assigned top similarity to more compact texts. Of them, the soft-cosine similarity model, using word2vec terms similarity, produced the most consistent results, for the human eye. Several of matched by this model pairs of paragraphs are listed below. The target/source text is identified by the title path: book title / section(chapter)title / [merged] paragraph index.

Before citing specific paragraphs, we outline some statistics. The source paragraphs have been matched against the target paragraphs in two regimes: 1. Matching vs a single paragraph containing the AoH message 12-points summary. 2. Matching vs a list of (merged) paragraphs of AoH message. In both cases top-64 by similarity pairs of matched paragraphs have been logged and reviewed. Also, to estimate the contrast of the model output, was logged the average similarity between a source and a target paragraph among all possible combinations. These statistics are listed in table 8.

What is notable in paragraph matches:

- When source paragraphs were matched vs all target paragraphs, a single paragraph from Good T. Alien Base, 1999 / Chapter 11. Let Humanity Beware! got listed 16 times among the 64 top rankings, matching vs

different target paragraphs. That paragraph is part of a message, which is said to be transmitted to humanity from an ET source in person via Bobby, a Philipino pianist, during his study in Vienna in 1962.

- When source paragraphs were matched vs the '12 points briefing summary' or vs all target paragraphs, 10 or 25 out of 64 top ranking source texts correspondingly came from the David Jacobs books, mainly 'Walking Among Us' and 'The Threat'. This confirmed anticipated high relevance of AoH message to the UFO abductions research results.

7.3.1 Paragraph match example 1

Similarity	0.535
Target title:	Summers M. Allies of Humanity, V4, 2019 / First Briefing: Who We Are and What We Represent / 001
Source title:	Good T. Alien Base, 1999 / Chapter 11. Let Humanity Beware! / 050

Target text (shortened):

... We are here to give you counsel and perspective and to correct many of your assumptions or beliefs that could jeopardize your ability to understand your situation and the great challenge you now face from Intervention in your world from beyond by other races who are here to take advantage of your weakness, your divisions, and your superstitions for their own purposes-to establish their precedence in this world without the use of force. We have spoken of this at great length in our previous Briefings. But now we must tell you about some other things that are important to know about the meaning of freedom in this universe-what it will mean for your world, the great change that it will bring to your understanding, and the greater unity it will require amongst your nations. For you will not remain a free world if you continue your struggle with one another-your dissensions, your competition, your violence, your wars. Fighting amongst yourselves, you have no idea of the presence of those who are intervening, or the many eyes in your local universe who are watching you, and all the errors of humanity being broadcast out into space for the discreet observer to see-your tribes warring with one another in a beautiful world that is desired by others. And though there are not many other races or nations that are aware of your existence, the challenge is yet very great, very real and growing all the time. So we must speak of things now that have to do with your future, that have to do with the possibility of humanity becoming a free race in a universe where competition has been carried to greater levels than you understand. We give this with great respect for humanity because you have kept your religious traditions alive, and the power of Knowledge within the individual is still recognized and honored here, where it has been lost and forgotten or never realized in so many other places. For the great technological nations in your local region of space are largely devoid of religion, spirituality and the higher ethics that you may still hold dear, which we value as well. You stand at the great turning

point of whether you will fall into submission to foreign powers or whether you will rise up from your endless conflicts to establish yourself as a free nation. ...

Source text:

Observe carefully the great mass of humanity killing each other through centuries of war and strife. And for what purpose? For power and domination; the intense desire to dominate and subjugate . There are thousands of good people on your planet, but the mean and selfish humanity outnumbers the good by millions and millions. Unless there is a radical change starting now, from an individual's mind, within a family and family relations, within a nation and between nations, your people will all be destroyed by their own selfishness for power and domination. Not until your humanity is completely wiped out will there ever be peace on Earth. The danger of atomic war between nations is imminent... It may not be now, it may be centuries from now, but the end will still come. There is no turning back for your people. Some day you will all be wiped out by your own greediness, and if a few good people live through, then they will propagate and breed an unselfish humanity and no longer will there be a continuous strife between nations, within a nation, within a family and family relations, and within an individual mind, and there will be peace on Earth at last. There is a great and possible danger, too, that your humanity's intense desire to conquer, eventually seeking power and domination over the other planets, will mean only a complete massacre for Earthmen, because other planets will retaliate with terrifying power and force, only because of their fear of your selfish humanity coming to their planets and spreading greed and evil around. This is our message. Transmit it and let humanity beware!

7.3.2 Paragraph match example 2

Similarity: 0.473

Target title: Summers M. Allies of Humanity, V1, 2001 / First Briefing: The Extraterrestrial Presence in the World Today / 034

Source title: Jacobs D. Walking Among Us, 2015 / 2. Abductees, Aliens, and the Program/031

Target text:

Now we must speak of the fourth area in which your visitors seek to establish themselves, and that is through interbreeding. They cannot live in your environment. They need your physical stamina. They need your natural affinity with the world. They need your reproductive abilities. They also want to bond with you because they understand that this creates allegiance. This, in a way, establishes their presence here because the offspring of such a program will have blood relations in the world and yet will have allegiance to the visitors. Perhaps this seems incredible, yet it is so very real. The visitors are not here to take your reproductive abilities away from you. They are here to establish themselves. They want humanity to believe in them and to serve them. They want humanity to work for them. They will promise anything, offer anything and do anything to achieve this goal. Yet though their persuasion is great, their numbers are small. But their influence is growing and their program of interbreeding,

which has been underway for several generations, will eventually be effective. There will be human beings of greater intelligence but who do not represent the human family. Such things are possible and have occurred countless times in the Greater Community. You have only to look at your own history to see the impact of cultures and races upon one another and to see how dominating and how influential these interactions can be. Thus, we bring with us important news, serious news. But you must take heart, for this is not a time for ambivalence. This is not a time to seek escape. This is not a time to concern yourself with your own happiness. This is a time to contribute to the world, to strengthen the human family and to call forth those natural abilities that exist in people—the ability to see, to know and to act in harmony with one another. These abilities can offset the influence that is being cast upon humanity at this time, but these abilities must grow and be shared. It is of the utmost importance.

Source text:

It is clear now that integration into human society is the aliens' primary goal. All aliens are part of a dedicated program to integrate hybrids into humanity. They perform their duties to fulfill the program's purposes and are all interconnected neurologically. All give allegiance to the insectalins. They do not talk about healing the Earth, or healing or enlightening humans, or joining together with humans in a cosmic community, or anything else that might help humans. They talk about sending aliens to Earth to live here undetected and eventually to supplant humans. Although the alien program is one of integration, exactly how The Change will take place is still a mystery. It is possible that integration into human society may merely be a preparation for The Change. If this is the case, then the program itself may be an ongoing integration into human society that will continue until the global population contains a critical mass of hybrids. In this case, The Change may be a final, as yet unknown, event or series of events that will culminate in the aliens' completion of their program. Whatever the final event is, it will be exclusively for the aliens' benefit.

7.3.3 Paragraph match example 3

Similarity: 0.381

Target title: Summers M. Allies of Humanity, 12 points briefing / Summary / 000

Source title: Hopkins B, Rainey C. Sight Unseen, 2003 / 26. Human Resources: Ours Or Theirs? / 036

Target text: <omitted>

Source text:

The old debate still rages on over what is the most powerful influence to shape human beings (or any sentient being): Is it nature or nurture? But it's possible that both our great human vulnerability and our strength lies in our ongoing desire to attend to both equally. We earthlings value our science; we don't mind tinkering with nature if it seems to confer some genetic advantage on us, rather than the next guy; we're even drawn to some of the more radical new abilities of genetics to make designer babies free of defects. But we know that nature won't

paint the whole picture, either-that these small bundles of neurons, muscles, blood, and bone, these offspring of ours, will also grow out of the environment we create and the examples we set. We will need to intentionally nurture our children into becoming loving, empathetic, and ethical adults, responsible to the rest of society. But what happens to a species like ours that might become infiltrated by a subpopulation of bioengineered people who never knew the nurturance side of the equation? ETs may be on the way to creating an increasingly perfect facsimile of a human being, but we don't yet see convincing evidence that they understand what lies behind the scenery of a fully fleshed-out human existence-qualities such as emotional openness, the courage to act independantly as a moral being, the desire to nurture others as you were once nurtured-or to nurture another person even if you were not. We don't know if this new variant species of mankind, intentionally created by the Other, is here to offer us earthlings fresh knowledge and profound insights from the farside of the cosmos, or whether it intends to deliberately betray us into the service of its own extraterrestrial creators. In the meantime, it seems that all we can do is to continue to bear witness and offer modest emotional support-through conversations, letters, phone calls, and writing-to the thousands of people undergoing these experiences. We, as well as the abductees, look toward the day that it dawns on the scientific, political, educational, and spiritual institutions of this small planet that their understanding and guidance are terribly lacking in dealing with this strange experience that touches every dimension of what it means to be human.

7.4 Detailed results

More detailed results are available in the project repository in the 'analysis' folder in the .json file format. These include:

- Detailed rankings of source texts to target texts for the book and section structure level. For book level, full rankings of similarity of source books to each target book are given. For section level, top-64 most similar source sections to each target section are given. The rankings are in the files with the name pattern: "ts_sim_<term-sim-model>_<level>_<frequency-mode>_closest.json", where term-sim-model, level and frequency-mode are as listed in 6.2.1; 'ts_sim' stands for 'target-to-source-similarity'.
- Top-64 pairs of most similar source and target (merged) paragraphs for the soft-cosine similarity models. These are in the files with the name pattern: "match_merged_pg[_12points]_<term-sim-model>_<frequency-mode>.json". The files with "_12points" match source paragraphs vs a single paragraph with the AoH message 12 points summary.

These files also include the similarities to aggregated source/target texts, and averaged similarity taken over all source-to-target text combinations; their special titles described in 6.2.

8 Technical conclusion

In this study was collected and marked up the internal structure of a unique dataset for comparing the texts of Allied-of-Humanity message, published by M. Summers, against a set of selected important UFO research books. The comparison was performed on 3 levels of text structure – between books, sections and (enlarged) paragraphs. As baselines for comparison task were used simple models of cosine similarity between documents, represented as bags-of-words- or TFIDF- vectors. Comparison, taking into account the terms’ similarity (provided by Word2Vec or FastText models, trained on the dataset corpus), was performed using soft-cosine similarity models. Different models, including simple baselines, demonstrated high relevance of AoH message texts to the following themes in the UFO research: the goals of the UFO abductions program; the long-term implications of the ongoing, though secret, contact with ET for humanity; spiritual/metaphysical dimension of UFO problem. Consistency between the outputs of different similarity models was shown via calculating Kendall correlation between rankings they produced.

The following areas of the present work need obvious improvements:

1. Further partitioning of the dataset texts into subsections. So far the subsections were marked up only for references/notes to exclude such text from analysis.
2. Uniting paragraphs to larger fragments: while the introduced simple algorithm did help to match larger and more meaningful text fragments, its results were often unsatisfactory: e.g. it did not unite several paragraphs, going as subsequent points/arguments for an encompassing statement. More advanced language models could be applied here. Or a different algorithm, enlarging fragments simultaneously with comparison.
3. The paragraph-to-paragraph matching revealed a need to refine the comparison task. The top-similar fragments sometimes (especially with FastText term similarity model) are texts which are really similar to each other in language, but are not very interesting by their content, e.g. being extracts from (self) dialogues. An obvious for human eye match between AoH’s point that the visitors seek on Earth biological resources, and a similar hypothesis raised by UFO researchers [26, Conclusion], did not get into top rank. What seems necessary here, is a kind of filter for the depth of the text content, so as to select for comparison only the fragments of high importance for society.

9 Final remarks

It is opinion of this author that the ET Intervention, described in AoH message, does have place, and that its activity can be discerned in the following areas:

1. Accelerated accumulation of the economical wealth in the hands of the few, often in the form of trusts/funds, which possess financial resources comparable to large states, are unaccountable to public, and attempt, with some success, to dictate global policies for the nation-states, companies and society. [59, 60]
2. Aggressive attack on the traditional family values, propaganda of homosexual marriages and transgenderism, promoted in many Western countries on the state level. The link to Intervention here might be twofold:
 - The homosexual couples do not reproduce naturally. If they do, it is with help of reproduction clinics, artificial insemination and surrogate mothers. Usage of reproduction technologies potentially opens the door for influencing the properties of progeny. This is in line with the AoH-described practice of Collectives breeding the offspring to serve a specific purpose (recall the “Brave New World” by Aldous Huxley) and with the description of producing alien-human hybrids in the UFO abduction research literature.
 - The artificial reproduction breaks the emotional bond between parents and children. Development of empathy, passing the intrinsic life experience becomes more difficult, especially given that in homosexual couples children face higher risks of sexual abuse. Breaking the parents-child link makes it easier to “reprogram” younger generations of society by propaganda disguised as entertainment and education.
3. The whole bunch of things pushed onto society by the World Economic Forum and its proponents, aka the Great Reset: deindustrialization under pretext of addressing the climate change – ESG policies; infringement of individual rights to move and to work under pretext of protecting people health; injection of nano-technological devices disguised as vaccines; the concept of trans-humanism; Central Bank Digital Currencies as programmable money, granting the governments total economical control over people; etc. [61].

In this context, it is worth pointing out that Allies-of-Humanity message is only part of a more broader message, transmitted to the world through Marshall Summers. The other part is accessible at www.newmessage.org and is said to include the methods of individual development of natural human capabilities, named Knowledge, which would be an essential tool to offset the Intervention. There are remarkable parallels between the description of Knowledge therein and the historical accounts of the human psychic abilities: mutual activation and relation to sexual energies among the first. We refer to [62] and to [63] for a thorough historical review of Psi studies.

References

- [1] Kilgariff, Adam (2001). Comparing Corpora. *International Journal of Corpus Linguistics*, 6:1 (2001), 97–133.
- [2] Gomaa. Wael H.; Fahmy, Aly A. (2013). A Survey of Text Similarity Approaches. *International Journal of Computer Applications (0975 – 8887)* Volume 68– No.13, April 2013.
- [3] Mihalcea, R., Corley, C., Strapparava, C. (2006). Corpus based and knowledge-based measures of text semantic similarity. In Proceedings of the American Association for Artificial Intelligence.(Boston, MA).
- [4] Mikolov T., Chen K., Corrado G., Dean J. (2013). Efficient estimation of word representations in vector space. *arXiv:1301.3781*.
- [5] Bojanowski P., Grave E., Joulin A., Mikolov T.(2016). Enriching Word Vectors with Subword Information. *arXiv:1607.04606*
- [6] Turney, P. (2001). Mining the web for synonyms: PMIIR versus LSA on TOEFL. In Proceedings of the Twelfth European Conference on Machine Learning (ECML).
- [7] Landauer, T.K., Dumais, S.T. (1997). A solution to plato’s problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge", *Psychological Review*, 104.
- [8] Matveeva, I., Levow, G., Farahat, A., Royer, C. (2005). Generalized latent semantic analysis for term representation. In Proc. of RANLP.
- [9] Sidorov G., Gelbukh A., Gómez-Adorno H., Pinto D. Soft Similarity and Soft Cosine Measure: Similarity of Features in Vector Space Model *Computación y Sistemas, vol. 18, núm. 3, julio-septiembre, 2014, pp. 491-504*.
- [10] Delphine Charlet and Geraldine Damnati. 2017. SimBow at SemEval-2017 Task 3: Soft-Cosine Semantic Similarity between Questions for Community Question Answering. In *Proc. of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*. ACL, Vancouver, Canada, 315–319. doi: 10.1865
- [11] Summers M. Allies of Humanity, 12 points briefing. www.alliesofhumanity.org
- [12] Summers M. Allies of Humanity, Volume 1, 2001.
- [13] Summers M. Allies of Humanity, Volume 2, 2005.
- [14] Summers M. Allies of Humanity, Volume 3, 2012.
- [15] Summers M. Allies of Humanity, Volume 4, 2019.

- [16] Catoe, Lynn E. UFOs and Related Subjects: An Annotated Bibliography, 1969.
- [17] Marrs J. Alien Agenda: investigating the extraterrestrial presence among us, 1997.
- [18] Dolan, Richard. UFOs for the 21st Century Mind, 2014.
- [19] Ruppelt, Edward. The report on unidentified flying objects, 1956.
- [20] Jacobs, David. The UFO Controversy in America, 1976.
- [21] Good, Timothy. Above Top Secret: the worldwide UFO cover-up, 1988.
- [22] Good, Timothy. Beyond Top Secret: The Worldwide UFO Security Threat, 1996.
- [23] Good, Timothy. Need to know: UFOs, the Military and Intelligence, 2007.
- [24] COMETA report. UFOs and Defense-What Should We Prepare For, 1999. English translation.
- [25] Hall, Richard. The UFO Evidence, Volume II, A 30-year report, 2000.
- [26] Dolan, Richard. UFOs and the National Security State: Chronology of a Cover-up: 1941-1973, 2002.
- [27] Dolan, Richard. UFOs and the National Security State: The Cover-Up Exposed: 1973-1991, 2009.
- [28] McDonald, James. Science in default: 22 years of inadequate UFO investigations, 1969.
- [29] Hynek, Allen J. The UFO Experience - A Scientific Inquiry, 1972.
- [30] Hill, Paul. Unconventional Flying Objects - a scientific analysis, 1995.
- [31] Friedman, Stanton. Flying Saucers and Science, 2008.
- [32] Hopkins, Budd. Missing Time: A Documented Study of UFO Abductions, 1981.
- [33] Hopkins, Budd. Intruders: the incredible visitations at Copley Woods, 1987.
- [34] Jacobs, David. Secret Life: Firsthand Accounts of UFO Abductions, 1993.
- [35] Jacobs, David. The Threat: Revealing the Secret Alien Agenda, 1998.
- [36] Jacobs, David. Walking Among Us: The Alien Plan to Control Humanity, 2015.
- [37] Mack, John E. Abduction-Human Encounters with Aliens, 1994.

- [38] Turner, Karla. Taken-Inside the Alien-Human Abduction Agenda, 1994.
- [39] Hopkins, Budd; Rainey, Carol. Sight Unseen: Science, UFO Invisibility and Transgenic Beings, 2003.
- [40] Clelland, Mike. The Messengers: Owls, Synchronicity and the UFO Abductee, 2015.
- [41] Leir, Roger. The Aliens and the Scalpel, 2005.
- [42] Howe, Linda. An Alien Harvest: Further Evidence Linking Animal Mutilations And Human Abductions To Alien Life Forms, 2014.
- [43] Wood, Robert; Redfern, Nick. Alien Viruses: Crashed UFOs, MJ - 12, & Biowarfare, 2013.
- [44] Crystall, Ellen. Silent Invasion, 1991.
- [45] Swann, Ingo. Penetration: The question of extraterrestrial and human telepathy, 1998.
- [46] Bergrun, Norman. The Ringmakers of Saturn, 1986.
- [47] Andrews, Colin. Crop Circles - Signs of Contact, 2003.
- [48] Alexander, Steve and Karen. Crop Circles Signs, Wonders and Mysteries, 2009.
- [49] Corso, Philipp. The Day After Roswell, 1997.
- [50] Good, Timothy. Alien Contact - Top-Secret UFO files revealed, 1991.
- [51] Good, Timothy. Alien Base - The Evidence for Extraterrestrial Colonization of Earth, 1999.
- [52] Good, Timothy. Unearthly Disclosure: Conflicting Interests in the Control of Extraterrestrial Intelligence, 2000.
- [53] Good, Timothy. Earth - the aliens enterprise, 2013.
- [54] Sauder, Richard. Underground Bases And Tunnels: What is the government trying to hide?, 2014.
- [55] Marler, David. Triangular UFOs: An Estimate of the Situation, 2013.
- [56] Bernstein, Carl. 1977. The CIA and the Media: How America's Most Powerful News Media Worked Hand in Glove with the Central Intelligence Agency. Rolling Stone, 20 October, 55-67.
- [57] Ulfkotte, Udo. Journalists for Hire: How the CIA Buys the News, 2017.
- [58] deHaven-Smith, Lance, prof. of political science. Conspiracy Theory in America, 2013.

- [59] Glattfelder J.B., Vitali S., Battiston S. The network of global corporate control. PLoS ONE 6(10), 2011.
- [60] Engdahl F. W. How Blackrock Investment Fund Triggered the Global Energy Crisis. www.globalresearch.ca, May 14, 2023.
- [61] The Top Ten Creepiest and Most Dystopian Things Pushed by the World Economic Forum (WEF), www.globalresearch.ca, October 17, 2022.
- [62] Swann, Ingo. Secrets of Power. Volume II. The vitalizing of individual powers, 2002.
- [63] Swann, Ingo. Psychic Sexuality: The Bio-Psychic "Anatomy" of Sexual Energies, 1998.

Title	Similarity
Model – Cosine Similarity, term similarity – none	
Dolan, 2014 / 1. What Are Ufos?	0.900
Dolan, 2014 / 12. Into The Future, Into Ourselves	0.900
Dolan, 2014 / 10. Who Are They? What Do They Want?	0.899
Marrs J. Alien Agenda, 1997 / 12. A Metaphysical Exam	0.886
Jacobs D. Walking Among Us, 2015 / 10. Integration And Speculation	0.883
Hopkins B, Rainey C. Sight Unseen, 2003 / 22. Behind The Scenery	0.880
Jacobs D. Walking Among Us, 2015 / 2. Abductees, Aliens, And The Program	0.879
Andrews C. 2003 / 13. Crop Circle Hoaxes And The Circus In The Fields	0.878
Dolan, 2014 / 11. Weird Science	0.876
Hopkins B, Rainey C. Sight Unseen, 2003 / 18. Vulnerable In A Thousand Ways	0.876
<SOURCE>	0.842
<SOURCE_MICRO>	0.754
Model – Soft Cosine Similarity, term similarity – word2vec	
Dolan, 2014 / 12. Into The Future, Into Ourselves	0.944
Friedman S. Flying Saucers And Science, 2008/ Conclusion: What Does It Matter?	0.942
Dolan, 2014 / 1. What Are Ufos?	0.936
Dolan, 2014 / 10. Who Are They? What Do They Want?	0.934
Dolan, 2014 / Introduction	0.925
Crystall E. Silent Invasion, 1991 / 14. The Human Factor	0.924
Jacobs D. Walking Among Us, 2015 / 10. Integration And Speculation	0.923
Andrews C. Crop Circles, 2003 / Closing Words: Carrying Out Instructions	0.920
Andrews C. Crop Circles, 2003 / 18. Half And Half: Where I Stand Today	0.917
Hopkins B, Rainey C. Sight Unseen, 2003 / 22. Behind The Scenery	0.917
<SOURCE>	0.874
<SOURCE_MICRO>	0.784
Model – Soft Cosine Similarity, term similarity – fasttext	
Dolan, 2014 / 12. Into The Future, Into Ourselves	0.925
Dolan, 2014 / 1. What Are Ufos?	0.916
Dolan, 2014 / 10. Who Are They? What Do They Want?	0.914
Friedman S. Flying Saucers And Science, 2008 / Conclusion: What Does It Matter?	0.910
Dolan, 2014 / Introduction	0.901
Hopkins B, Rainey C. Sight Unseen, 2003 / 22. Behind The Scenery	0.895
Friedman S. Flying Saucers And Science, 2008 / 6. The Ufo 'Why' Questions	0.894
Marrs J. Alien Agenda, 1997 / 12. A Metaphysical Exam	0.894
Andrews C. Crop Circles, 2003 / 13. Crop Circle Hoaxes And The Circus In The Fields	0.891
Leir R. The Aliens And The Scalpel, 2005 / 15. Into The Future	0.889
<SOURCE>	0.850
<SOURCE_MICRO>	0.751

Table 5: The rankings of similarity of SOURCE **section**-level texts to the aggregated TARGET text; documents – **bag-of-words** vectors, term similarity – none, word2vec, fasttext.

Title	Similarity
Model – Cosine Similarity, term similarity – none	
Dolan R., 2014 / 12. Into The Future, Into Ourselves	0.272
Dolan R., 2014 / 1. What Are Ufos?	0.226
<SOURCE>	0.219
Dolan R., 2014 / 10. Who Are They? What Do They Want?	0.184
Good T. Alien Base, 1999 / 4. Steps To The Stars	0.175
Jacobs D. The Threat, 1998 / 12. Life As We Know It?	0.172
Marrs J. Alien Agenda, 1997 / 12. A Metaphysical Exam	0.172
Mack J. Abduction, 1994 / 16. Alien Intervention And Human Evolution	0.171
Hall R. The Ufo Evidence, V2, 2000 / XVI. Implications And Issues	0.169
Dolan R. Ufos And The National Security State-V2, 2009 / Conclusion	0.169
Dolan R., 2014 / 8. Contact And Abduction	0.163
<SOURCE_MICRO>	0.053
Model – Soft Cosine Similarity, term similarity – word2vec	
Dolan R., 2014 / 12. Into The Future, Into Ourselves	0.683
Marrs J. Alien Agenda, 1997 / Epilogue. New Realities	0.634
Dolan R., 2014 / 1. What Are Ufos?	0.616
Dolan R. Ufos And The National Security State-V2, 2009 / Conclusion	0.590
Mack J. Abduction-Human Encounters With Aliens, 1994 / 16. Alien Intervention And Human Evolution	0.578
Friedman S. Flying Saucers And Science, 2008 / Conclusion: What Does It Matter?	0.576
Dolan R., 2014 / 10. Who Are They? What Do They Want?	0.573
Cometa Report, 1999 / 13. Political And Religious Implications	0.561
Hall R. The Ufo Evidence, V2, 2000 / XVI. Implications And Issues	0.556
Jacobs D. The Threat, 1998 / 13. Accepting The Unacceptable	0.549
<SOURCE>	0.418
<SOURCE_MICRO>	0.224
Model – Soft Cosine Similarity, term similarity – fasttext	
Dolan R., 2014 / 12. Into The Future, Into Ourselves	0.639
Dolan R., 2014 / 1. What Are Ufos?	0.575
Dolan R. Ufos And The National Security State-V2, 2009 / Conclusion	0.529
Dolan R., 2014 / 10. Who Are They? What Do They Want?	0.525
Marrs J. Alien Agenda, 1997 / Epilogue. New Realities	0.523
Mack J. Abduction-Human Encounters With Aliens, 1994 / 16. Alien Intervention And Human Evolution	0.522
Marrs J. Alien Agenda, 1997 / 12. A Metaphysical Exam	0.515
Hall R. The Ufo Evidence, V2, 2000 / XVI. Implications And Issues	0.512
Dolan R. Ufos And The National Security State-V2, 2009 / Introduction	0.505
Cometa Report, 1999 / 13. Political And Religious Implications	0.497
<SOURCE>	0.441
<SOURCE_MICRO>	0.227

Table 6: The rankings of similarity of SOURCE **section**-level texts to the aggregated TARGET text; documents – **TFIDF** vectors, term similarity – none, word2vec, fasttext.

Model 1/2	(w2v,bow)	(ft,bow)	(none,tfidf)	(w2v,tfidf)	(ft,tfidf)
(none,bow)	0.46, 4e-6	0.48, 9.0e-7	0.27, 0.02	0.35, 1.3e-3	0.36, 1.2e-3
(w2v,bow)		0.72, 2.9e-16	0.18, 0.1	0.38, 2.1e-4	0.33, 2.0e-3
(ft,bow)			0.20, 6.5e-2	0.36, 7.7e-4	0.34, 1.5e-3
(none,tfidf)				0.38, 2.4e-4	0.48, 1.2e-6
(w2v,tfidf)					0.66, 5.4e-12

Table 7: The averaged Kendall-tau correlation between top-64 rankings of section-level models. Each row/column represents a similarity model, the cell of row/column intersection – results of comparison of rankings of 2 models. In each cell the 1st number is the arithmetic average of Kendall correlation coefficient, the 2nd number is the geometric average of associated p-value. Models’ abbreviations: none – cosine similarity baseline model; w2v/ft – soft-cosine similarity model using word2vec/fasttext terms similarity. bow/tfidf - frequency mode used for the document vectors.

Matching source paragraphs vs:	top-64 matches	avg. source-to-target	avg. source-to - (united target)
Single paragraph with 12-points summary	0.317-0.431	0.0835	0.0835
List of all target’s paragraphs	0.470-0.565	0.0737	0.122

Table 8: The average and range of top-64 ranking similarity between source and target (merged) paragraphs for soft-cosine similarity model with term similarity – Word2Vec, frequency mode – TFIDF.