This is a peer-reviewed, author's accepted manuscript of the following conference proceedings chapter: Pinkosova, Z., McGeown, W. J., & Moshfeghi, Y. (2023). Revisiting neurological aspects of relevance: an EEG study. In G. Nicosia, G. Giuffrida, V. Ojha, E. La Malfa, G. La Malfa, P. Pardalos, G. Di Fatta, & R. Umeton (Eds.), Machine Learning, Optimization, and Data Science - 8th International Conference, LOD 2022, Revised Selected Papers (pp. 549-563). (Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Vol. 13811 LNCS). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-031-25891-6_41

Revisiting Neurological Aspects of Relevance: An EEG Study

Zuzana Pinkosova[1][0000−0002−6238−9264], William J. McGeown[2][0000−0001−7943−5901], and Yashar Moshfeghi[1][0000−0003−4186−1088]

[1] NeuraSearch Laboratory, University of Strathclyde, Glasgow, UK
zuzana.pinkosova@strath.ac.uk, yashar.moshfeghi@strath.ac.uk

[2] Neuroanalytics Laboratory, University of Strathclyde, Glasgow, UK
william.mcgeown@strath.ac.uk

Abstract. Relevance is a key topic in Information Retrieval (IR). It indicates how well the information retrieved by the search engine meets the user's information need (IN). Despite research advances in the past decades, the use of brain imaging techniques to investigate complex cognitive processes underpinning relevance is relatively recent, yet has provided valuable insight to better understanding this complex human notion. However, past electrophysiological studies have mainly employed an event-related potential (ERP) component-driven approach. While this approach is effective in exploring known phenomena, it might overlook the key cognitive aspects that significantly contribute to unexplored and complex cognitive processes such as relevance assessment formation. This paper, therefore, aims to study the relevance assessment phenomena using a data-driven approach. To do so, we measured the neural activity of twenty-five participants using electroencephalography (EEG). In particular, the neural activity was recorded in response to participants' binary relevance assessment (relevant vs. non-relevant) within the context of a Question Answering (Q/A) Task. We found significant variation associated with the user's subjective assessment of relevant and non-relevant information within the EEG signals associated with P300/CPP, N400 and, LPC components, which confirms the findings of previous studies. Additionally, the data-driven approach revealed neural differences associated with the previously not reported P100 component, which might play important role in early selective attention and working memory modulation. Our findings are an important step towards a better understanding of the cognitive mechanisms involved in relevance assessment and more effective IR systems.

Keywords: Information Retrieval · Relevance Assessment · Binary Relevance · Brain Signals · EEG · ERPs · Cognitive Processes

1 Introduction

Relevance assessment plays a central role in Information Retrieval (IR), denoting how well the document retrieved by an IR system meets the searcher's information need (IN) submitted as a query to the system (see Figure 1). Although IR covers documents containing different modalities (e.g. videos or images) the most information consumption happens in textual format [3]. Assessing relevance of textual documents, given IN, involves several cognitive processes including reading comprehension. Therefore, it is one of the most complex cognitive activities in IR [13]. In addition, despite recent findings supporting the idea of categorical thinking [38], relevance assessment has been primarily investigated in binary terms (i.e. content judged as 'relevant' vs 'non-relevant') [42]. Therefore, this work aims to focus on textual binary relevance assessment, which would enable us to compare experimental results obtained using a data-driven approach with previously reported results associated with textual binary relevance assessment formation.

Fig. 1: The interactive IR components.

Recently, NeuraSearch research, which is user-centred research bridging neuroscience and IR [29], has significantly contributed to our understanding of relevance assessment and associated cognitive processes. However, past brain-imaging studies investigating relevance assessment have either misplaced relevance within the context of word-relatedness (i.e. IN of the user was not considered) [14] or (when considering IN) predominantly focused on ERP component-driven analysis [38]. Although investigating relevance assessment employing the component-driven approach provides invaluable insights, its ability to detect and quantify other previously not reported potential components that could arise from this phenomenon is limited. This work, in contrast, employs a data-driven approach to avoid the potential analytical bias introduced by the restriction to distinct ERPs reported by previous studies [44]. This is the first NeuraSearch EEG study investigating binary relevance assessment using a data-driven approach to gain a holistic view of ERP components underpinning complex cognitive processes associated with relevance assessment.

2 Related Work

Relevance Assessment. Relevance has been considered in IR as a multi-dimensional [42], dynamic and complex phenomenon [8]. Previous studies have shown that relevance is difficult to quantify and depends on the users' perception of information relating to the specific IN situation at a certain time point [42]. Therefore, relevance assessment strongly depends on the user's internal cognitive states. In terms of relevance granularity, binary division has been the prevalent approach in the field, keeping the assessment cost low while maximising the number of relevant documents per topic, guaranteeing measure stability [42, 46][3]. In this work, thus, we capture the users' binary relevance assessment in real-time as they engage in the question-answering task.
Neurasearch Research. Neuroscience has significantly contributed to the understanding of IN [34], query formulation [17], search [31] and relevance assessment (e.g. [20, 38]) leading to the potential development of novel information search models [31] that can incorporate user's neurophysiological responses to the presented information. The most frequently used neuroimaging methods in the IR have been functional magnetic resonance (fMRI) [30–33, 36, 37], magnetoencephalography (MEG) [20] and electroencephalography (EEG) [2, 17, 22]. Within the relevance assessment context, the above-mentioned brain imaging techniques have been employed to investigate different aspects of neurological relevance assessment, such as the brain's functional connectivity [31], underlying cognitive processes and their timing [2].
NeuraSearch and Relevance Assessment. The neuroscientific approach might be categorised in two ways based on the experimental design used to investigate relevance assessment. The first line of brain-imaging research (NeuraSearch) has considered users' IN and positioned relevance assessment within the IR task. Moshfeghi and colleagues [30] used fMRI to localise cortical activity differences during the processing of relevant vs non-relevant images that were related to visuospatial working memory [31, 34]. Relevance assessment has also been studied within the context of the IR task for stimuli of different modalities, such as text [17], images [2] and videos [22]. Allegretti et al. examined the processing of relevant vs non-relevant images, finding the most significant differences to occur between 500 − 800ms, reaching the peak in the central scalp areas [2]. Kim and Kim [22] explored the topical relevance of video skims and classified the data based on two specific ERP components (N400 and P600), which are indicators of relevant and non-relevant assessments. Furthermore, recent findings have shown that relevance assessment can be automatically predicted using EEG data while the user engages with the IR task [17]. The results of these studies suggest that human mental experience during the relevance assessment can be understood and accurately decoded.

Another approach employing EEG has placed relevance assessments in the word associations context [14, 15]. Within the task, participants did not experi-

[3] Recent studies have discussed higher relevance granularity [38, 52], however, it is out of the scope of this paper.

ence IN, but they judged associations between words and topics. The study findings have shown that neurological signals differ when subjects process relevant vs non-relevant words [15]. Later, Eugster et al. [14] introduced a brain-relevance paradigm enabling information recommendation to users without any explicit user interaction, based on EEG signals alone evoked by users' text engagement.

Data-driven Approach. Relevance assessment is a holistic cognitive response with underlying neuropsychological mechanisms that form more basic perceptional and cognitive features of some sort. While past feature classification and theory-driven approaches have contributed to the understanding of relevance assessment formation, these approaches can constrain knowledge by potentially overlooking all the possible features or dimensions that synthesise complex phenomena such as relevance assessment. On the other hand, a data-driven approach is a useful tool when it comes to making sense of complex behavioural responses during complex tasks. Despite its advantages, the data-driven analysis might be challenging because the EEG data frequently exhibit spatial heterogeneity, have non-stationary and multiscale dynamics, and are typified by substantial individual variability [12]. Many scientific observations of brain activity may be scope limited, constrained by opportunistic participant sampling, and have reduced reproducible controls. Relying on varied background assumptions while employing a data-driven exploratory approach might help to overcome the above-mentioned limitations and provide significant benefits associated with the potential discovery of novel, previously not reported cognitive phenomena.

3 Methodology

Participants. The study was carried out with a sample of twenty-five individuals who reported themselves to be neurologically and physically healthy. Seven participants were excluded from the analysis due to the high number of physiological artefacts present in the EEG data. The eighteen remaining participants (11 females and 7 males) were between 19 to 39 years old and with the mean age of 24.5 and the standard deviation (SD) 4.91 years. Participants were either native English speakers (8) or had high English proficiency. On average, participants indicated using search engines on average several times a day.

Design. This study followed a within-subject design in which participants engaged in the Q/A task. The independent variable was the user's binary relevance assessment (with two levels: "Non-Relevant" ('nr') and "Relevant" ('rel')). The dependent variable was the EEG signal gathered during the answer presentation. We controlled the number of relevant and non-relevant answers presented to the participant, but we did not control the number of words each participant saw. This allowed us to simulate an information search and retrieval, as participants were not required to read through the whole answer. Instead, they were able to terminate the answer presentation once the relevance assessment was made.

Stimuli Presentation. The stimuli were presented on a 22-inch colour Mitsubishi Diamond Pro 2040u CRT monitor using E-Prime 2.0. Participants were seated approximately 60 cm from the monitor, and response keys were located on a QWERTY keyboard. Text events were presented in Arial font, size 16.

Questionnaires. Throughout the experiment, participants filled in the Entry and Post-Task Questionnaires. The Entry questionnaire was administered at the beginning of the experiment to gather participants' demographic and background information. All participants were screened for eligibility to take part in the study. Any participants with a pre-existing neurological condition that might impact the EEG recordings were excluded. After completing the task, participants filled in the Post-Task questionnaire, assessing their task perception.

Question-Answering Data Sets. The data set employed in the study was developed and used by Moshfeghi, Triantafillou, and Pollick [34]. We have chosen this data set as it has been proven effective in investigating IR phenomena from a neuroscience standpoint. The data set was further adapted to address our research questions and expanded through the additional question and answer selection from TREC-8 and TREC-2001. These two Tracks were selected as they (i) cover a wide discipline range, (ii) they are independent of one another, and (iii) they provide a correct question answer. We ensured that the selected questions were accurate and not time-dependent or ambiguous. The data set contained 128 questions, answers, and relevance assessments in total. To reduce the fatigue, 64 questions were carefully selected for each participant to be balanced based on relevance (i.e. 50% relevant and 50% non-relevant), length (i.e. 50% long vs 50% short) and difficulty (i.e. 50% difficult vs 50% easy)[4]. This was done to reduce any possible bias that could result from the focus on one form of Q/A. An example of an easy question presented to the participants was "What is epilepsy?", which was followed by the short, relevant answer "Epilepsy is a brain disorder characterised by seizures". The order of the presented questions was randomised also for each participant.

EEG Recordings. Brain signals were acquired using 128-channel HydroCel Geodesic Sensor Nets (Electrical Geodesic Inc) and recorded within the standard EGI package Net Station 4.3.1. A Net Amps 200 amplifier was used for the recording and to facilitate the synchronisation between the behavioural response of the participant and their brain signals. To set the system for recording, we followed Electrical Geodesic Inc guidelines. We aimed to keep the electrode impedances below 50 kΩ, according to the recommended system value. Raw EEG data were recorded at a sampling rate of 1000 Hz and referenced to the vertex electrode (Cz).

Procedure. The experiment was carried out in accordance with the University of Strathclyde Ethics Committee guidelines. All participants were briefed and confirmed their willingness to voluntarily participate in the experiment. After that, they filled in an Entry Questionnaire. Prior to the main experimental trials, participants underwent a number of training trials, which resembled the main experimental task. Participants were able to repeat the training until they confirmed to have a good understanding of the procedure. In total, every participant

[4] To assess the difficulty level, two annotators separately judged question difficulty (i.e. difficult vs easy). The overall inter-annotator agreement was reasonably high (Cohen's kappa, $\kappa = 0.72$)

completed 64 trials. To avoid fatigue, the trials were split into two equally long blocks separated by a break. On average, participants were presented with 810.06 words ($\pm$ 134.77) and the main experimental task lasted approximately 53.69 minutes $\pm$ 9.74). After completing the main experimental task, participants were instructed to fill out the Post-Task Questionnaire.

Task. The schematic task representation is depicted in Figure 2. At the beginning of the task, participants were presented with the instructions. Next, they viewed a randomly selected question from the data set, followed by the fixation cross, which indicated the location of the answer presentation. To control free-viewing and minimise the presence of any confounding artefacts (i.e. saccades), the answer was presented in the middle of the screen word by word. Each word was presented for 950ms, which has been deemed to be a sufficient duration to model fluent reading and to avoid the overlapping effect of two consecutive words on the ERPs [14]. The ERP components were, therefore, time-locked to the word presentation. This approach has been commonly applied to examine neurological signatures of reading in the ERP studies (e.g. [11]).

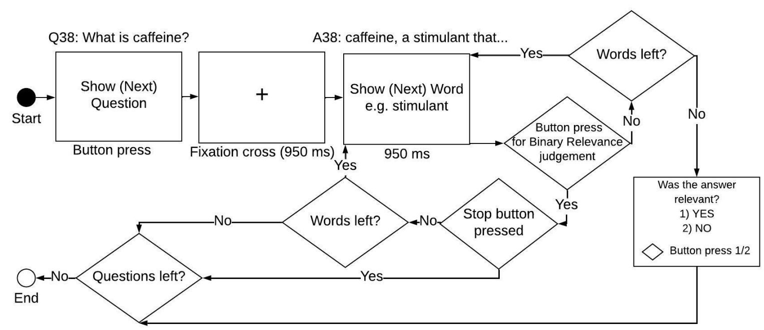

Fig. 2: The flow diagram of the task.

Participants were instructed to carefully read individual words that would form either relevant or non-relevant answers. Once participants gathered enough information, they had an option to terminate the word presentation sequence (and to continue to the next step), or to view the sequence in full. As brain activity was recorded during the reading, to avoid the possibility of confounding hemispheric effects (due to motor planning or execution), counter-balancing was used, and participants were instructed to interact with the keyboard using either their left or right hand. The interpretation of binary relevance categories depended on each participant's subjective understanding, which enabled capturing the subjective nature of relevance assessment [42].

Pilot Studies. Before commencing the main user study, we performed a pilot study with four participants whose data were not included in the final analysis.

Based on participants' feedback, it was determined that the participants were able to complete the study without problems.

Pre-processing Steps. The brain activity was recorded from participants as they engaged with relevant and non-relevant content, up to the point where the participant stopped the answer presentation. To prepare data for analysis, an automated pre-processing pipeline was built using EEGLAB and its associated toolboxes [10]. The EEG data pre-processing steps were based on Makoto's Pre-processing Pipeline[5]. All EEG data were first visually inspected. Then a low-pass filter of 30Hz was applied. We down-sampled the data from 1000Hz to 250Hz. Downsampling, a commonly applied procedure, is used to reduce file size for easier data manipulation. Then a high pass filter of 0.3Hz was applied. Filtering is another common procedure used to attenuate frequencies associated with noise rather than a signal of interest. We then automatically rejected bad channels (EEG sensors that were not functioning properly during the data acquisition and that were high in noise throughout the task. On average, we removed 13.94 bad channels ($\pm$7.67). The re-referencing to average (across all electrodes) was subsequently performed (to provide an approximation of zero microvolts for the reference at each timepoint). The CleanLine EEGLAB plugin was used to filter line noise. All epochs (the time windows of interest) were then extracted from 200ms before stimulus presentation to 950ms afterwards. To detect and remove components associated with ocular, cardiac and muscular artefacts based on their power spectrum and time-course, we performed Independent Component Analysis and then used ADJUST [28] to automatically reject components. A mean number of 18.17 ($\pm$9.17) components were removed. Bad channels were interpolated (reconstructed) using a spherical interpolation method. Next, we removed the two outermost belts of electrodes of the sensor net[6] which are prone to show muscular artefacts, following the approaches of Bian et al. [4] and Calbi et al. [6]. Epochs were then extracted again from 100ms before stimulus presentation to 950ms afterwards based on the stimulus labels for every condition of interest (i.e. 'rel' and 'nr'). We have used automatic epoch rejection based on thresholding (i.e. rejecting epochs by detecting outlier values greater than $\pm 100\mu V$). The mean number and SD of accepted and rejected epochs are displayed in Table 1. All epochs were baseline corrected. After pre-processing the data, we calculated the grand average for epochs of interest.

	Accepted Epochs		Rejected Epochs	
Condition	Mean	SD	Mean	SD
rel	399.11	88.01	69.39	89.55
nr	410.94	95.99	76.56	90.76

Table 1: The Mean number and SD of rejected epochs for 'rel' and 'nr' conditions

[5] https://sccn.ucsd.edu/wiki/Makoto's$_{p}reprocessing_pipeline$

[6] We removed 38 peripheral channels: E1, E8, E14, E17, E21, E25, E32, E38, E43, E44, E48, E49, E56, E57, E63, E64, E68, E69, E73, E74, E81, E82, E88, E89, E94, E95, E99, E100, E107, E113, E114, E119, E120, E121, E125, E126, E127, E128

Statistical Analysis of EEG data. Participants' brain activity was recorded while they engaged in relevant and non-relevant assessments (i.e. 'rel' vs 'nr') within the Q/A task. After data pre-processing, 49.65 % of accepted trials were marked as 'rel' and 50.35 % as 'nr'. To test whether there are statistically significant differences in the neurological processing associated with the judgement of 'rel' vs 'nr' information, we employed a data-driven approach, which is particularly effective in whole-brain analysis of complex mental phenomena as it minimises the upfront assumptions and allows for the contribution of many distinct areas at different time points [7, 27]. To identify significant cortical differences, we compared the values for 109 electrode pairs at every time point (every 4ms, 237-time points in total) over the 100 - 950ms time window. The initial time interval (0 - 100ms) was excluded from the main analysis as we were not interested in the initial sensory processing of stimulus features [25]. The data-driven approach applied a non-parametric permutation-based paired t-test (1000 permutations) using the *statcond* function implemented in the EEGLAB toolbox [10]. Differences were considered significant at a threshold of $p < 0.05$.
ROIs. As this study uses a data-driven approach, for optimal effect detection, the Regions of Interest (ROIs)[7] were determined based on statistically significant differences between compared conditions of interest. Therefore, we have used features of the data under analysis to position the ROIs. We were not interested in isolated electrodes where a test statistic might happen to be large. Instead, we applied the method used by Laganaro and colleagues [23]. To identify potential ROIs, we have only considered clusters with at least five electrodes next to each other extending over at least 20ms and retained with an alpha criterion of 0.05 [23].

4 Results

Questionnaire Results. Before analysing the main results, we measured participants' task perception using the Post-task Questionnaire using a 7-point Likert Scale (answers: 1: "Strongly Disagree", 2: "Disagree", 3: "Somewhat Disagree", 4: "Neither Agree or Disagree", 5: "Somewhat Agree", 6: "Agree", 7: "Strongly Agree"). Overall, the Post-task Questionnaire results suggest that participants did not perceive difficulties associated with the experimental design that might have caused them discomfort and impacted their engagement (Figure 3).

Binary Relevance. As described above, we used a data-driven analysis to investigate the brain activity differences during the 'rel' vs 'nr' content processing and this section presents the main experimental results.
100 - 200ms. The earliest differences in neural activity for the comparison 'rel' vs 'nr' conditions emerged in the 100 - 200ms interval. The 'rel' condition was associated with a significantly greater positivity in the right centro-parietal region and significantly greater negativity in the left fronto-central region compared to

[7] Region of Interest refers to a selected region of neighbouring electrodes that jointly and significantly contribute towards neurophysiological phenomena of interest

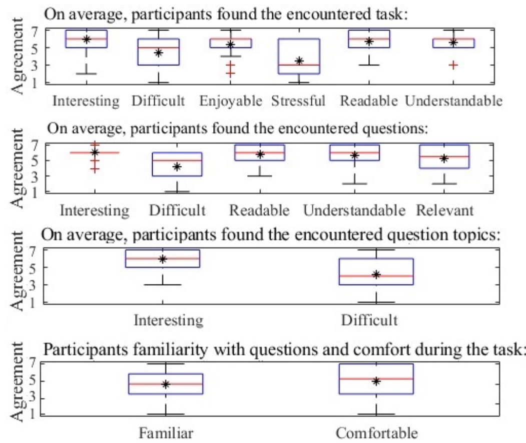

Fig. 3: Post-Task Questionnaire Results: The asterisk (*) denotes the mean values and the dot (·) denotes the outlier values.

the 'nr' condition. Significant electrode clusters, time intervals, and ERP waveforms, as well as topographic plots, are displayed in Figure 4, row I. Given the topographies and waveform peaks, the differences are likely to reflect variability in the P100 ERP component (similiar distributions reported, e.g., by [24]). The P100 ERP component reflects initial visual field activation and enhanced P100 amplitude observed during the processing of relevant information might suggest early selective attention allocation, with greater early attention allocated to relevant stimuli [26]. This early P100 selective stimulus encoding might affect later ERP components associated with working-memory [41], such as the P300 and LPC commonly reported in the relevance assessment studies [38].

250 - 300ms. The comparison of 'rel' and 'nr' conditions was associated with statistically significant differences in the right centro-parietal region within the 250 - 300ms interval (see Figure 4, row II). Closer inspection of the topographies revealed that the above effects were driven by greater positivity of the P300/Centro-parietal positivity (CPP) component across the 'rel' condition compared to the 'nr' condition. Obtained topographies with centro-parietal positivity are very similar to the P300 and CPP signal, peaking at around 300ms post-stimulus [48]. Both the CPP and P300 ERPs share many characteristics, such as dynamics and topography. While the P300 amplitude reflects attentional allocation [19] and task relevance [16], the CPP amplitude modulations are related to evidence accumulation during decision formation [48]. Significant differences might signal difference in effort for participants to process 'rel' and 'nr' information, which may induce differences in working memory load [50].

400 - 600ms. The processing of 'nr' compared to 'rel' content was associated with lower amplitude in an electrode cluster that bridged the right centro-parieto-temporal negativity within the 400 - 600ms time interval, as displayed in Figure 4, row III. The significant differences were driven by the higher centro-

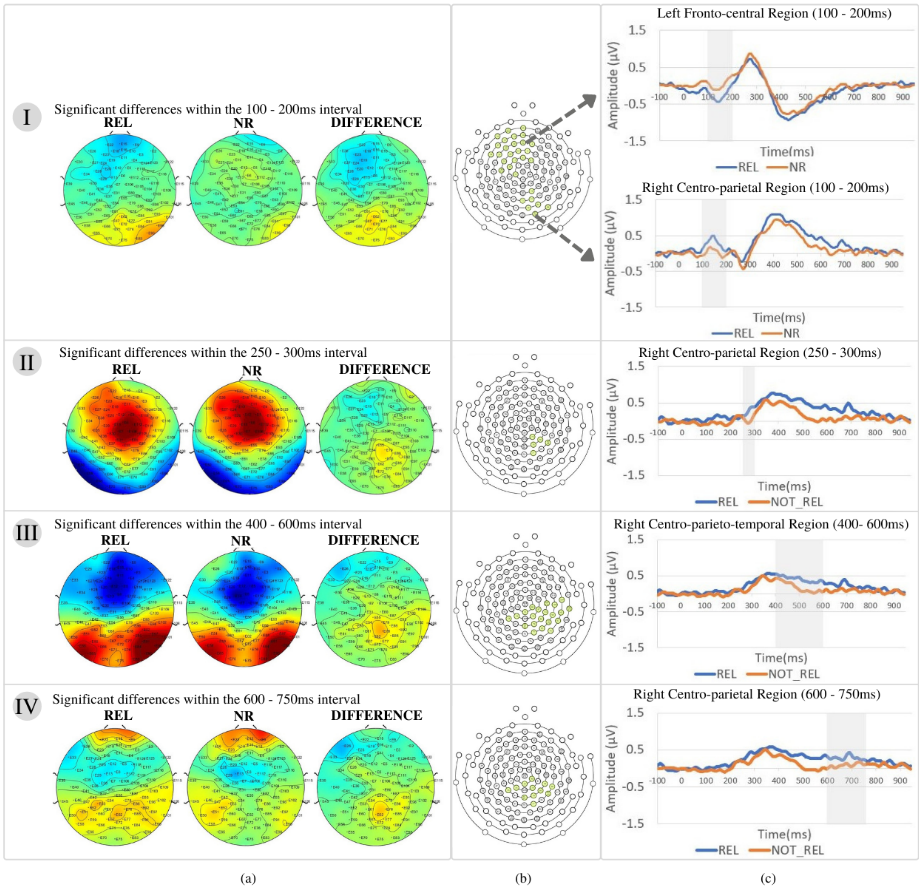

Fig. 4: (a) Topographic plots for 'rel' vs 'nr' conditions, including a mean difference plot for the 100 - 200ms (I), 250 - 300ms (II), 400 - 600ms (III) and 600 - 750ms (IV) time windows. Reddish colours of the scalp topography indicate positive ERP values, whereas bluish colours indicate negative ERP values. (b) The 128-channel net graph with highlighted statistically significant electrode sites for each significant time interval. (c) The comparison of grand averaged ERP waveforms for 'rel' (blue) vs 'nr' (orange) condition. Significant time intervals are highlighted in grey for each significant time period.

parieto-temporal negativity associated with 'nr' compared to the 'rel' condition. Observed anterior negativity and co-occurring posterior positivity reflect the N400 ERP component, with similar topographic distributions reported by previous studies [43, 47]. Lower N400 amplitudes in response to 'nr' content might be linked to the higher semantic incongruity between the question context and the given answer. That 'nr' content is linked with less positivity in this time window bears similarity to the results of e.g. [22], who found that irrelevant content produced more negative N400 responses.

600 - 750ms. The posterior positivity observed in the topographic plots within the 600 - 750ms time frame (see Figure 4, row IV) can be attributed to the LPC (Late Positive Component)[8] (e.g. [14, 15]). The LPC component is a positive-going deflection, emerging around 600ms post-stimulus, usually largest over the posterio-medial brain areas [9]. The positivity was significantly greater during the processing of 'rel' compared to the 'nr' content. Higher LPC amplitudes are associated with memory processing and decision-making, which might reflect effort invested in retaining relevant information during cumulative information exposure [22, 51]. Furthermore, higher LPC amplitude deflection in response to task-related stimuli have previously been reported [5], which is consistent with our findings.

5 Discussion & Conclusion

The current experiment was carried out to investigate binary relevance assessment using a data-driven approach to gain a better understanding of complex cognitive processes underpinning relevance assessment formation. In particular, the user's neural signals associated with 'rel' and 'nr' relevance assessments were recorded in real-time during a Q/A relevance assessment task. Using a data-driven approach we compared information judged as relevant to that which was not relevant. We found significant differences in neural activity associated with the user's perceived relevance in distinct significant time intervals linked to differences in topographies and ERP waveform distributions. Along with previously reported P300/CPP, N400 and LPC ERP components in studies investigating binary relevance (e.g., [22]), we have also observed significant differences in neural activity in the early time interval associated with the P100 component. Differences in the P100 component might signal early stages of selective attention allocation, indicating enhanced attention to 'rel' information during early sensory facilitation, which is then passed along to higher levels of cognitive processing.

P100. The 100 - 200ms time interval seems to be an early time point associated with the P100 ERP component, reflecting the participant's selective attention

[8] The terms LPC and P600 are commonly interchanged. Relevance assessment has frequently been linked to the P600 ERP component (e.g. [14]. However, the P600 component is mainly associated with 'syntactic re-analyses' in language studies. Therefore, the label LPC might be more appropriate to use while focusing on relevance assessment, as the LPC has been linked to memory and recognition processes.

modulation of task-relevant information. Past studies have provided a direct correlative link between the P100 and working memory performance, which might suggest that the P100 reflects participants' initial ability and processing effort to recognise relevant stimuli during relevance assessment formation [41]. Further studies could explore the relationship between the P100, P300 and LPC ERP components, related to attention during relevance assessment formation.

P300 / CPP. Observed P300/CPP topographies and ERP waveforms might indicate the allocation of greater attentional resources to information perceived as relevant. Relevant information might also be easier to process in terms of memory access and retrieval, leading to reduced cognitive load [1, 39]. The CPP component is equivalent to the P300 and previous research indicates that both components represent decision formation, through the information accumulation process, determining behaviour via a boundary-crossing criterion [35]. In addition, the P300/CPP amplitude is proportional to attentional engagement [18, 21], quality of useful information [19, 21], and the quality of information transmitted [21, 40]. Our results are consistent with previous studies suggesting that the P300 amplitude is higher for the target ('rel') than for non-target ('nr') stimuli [49].

N400. Significantly greater amplitude was observed for the 'rel' compared to the 'nr' condition, as indicated through the differences observed within the N400 topographies and time-window (see Figure 4, line III). The N400 component has been extensively researched in the sense of semantic processing. Past studies suggest that the N400 represents aspects of semantic information retrieval and integration [11]. N400 negativity is amplified by the processing of semantic mismatch (see e.g., [45]). Less negative deflections (indicated through the higher amplitude) for the 'rel' condition might, therefore, reflect higher answer relatedness to the question, which has been previously reported by [38], who found a reduced N400 for highly relevant words.

LPC. The amplitudes associated with the LPC component are significantly greater for content assessed as 'rel' compared to 'nr', as indicated through the differences observed within the LPC topographies (see Figure 4, line IV). The LPC is commonly reported to follow the N400 component and it is related to cumulative evidence exposure during decision-dependent tasks [51]. Therefore, the LPC amplitudes appear to be modulated by the participant's response to a stimulus, when the memory judgement at hand requires consideration of the relevant dimension in search tasks [51]. Furthermore, higher amplitudes recorded during the 'rel' condition might suggest the ease of the word categorisation process.

To conclude, our findings further our understanding of the concept of relevance and provide the evidence needed to strengthen its theoretical foundations. Overall, our results confirmed the empirical findings of previous studies examining textual relevance processing associated with the P300/CPP, N400 and LPC components. Additionally, the data-driven approach revealed neural differences in an early P100 component, which is a novel, previously not reported finding. Finally, we believe our conclusions constitute an important step in unravelling the nature of relevance assessment in terms of its electrophysiological modulation and operationalising it for the IR process.

Acknowledgement: This work was supported by the Engineering and Physical Sciences Research Council [grant number EP/R513349/1].

References

1. Ahmed, L., de Fockert, J.W.: Working memory load can both improve and impair selective attention: evidence from the navon paradigm. Attention, Perception, & Psychophysics **74**(7), 1397–1405 (2012)
2. Allegretti, M., Moshfeghi, Y., Hadjigeorgieva, M., Pollick, F.E., Jose, J.M., Pasi, G.: When relevance judgement is happening? an eeg-based study. In: SIGIR '15. p. 719–722. ACM, NY, USA (2015)
3. Barral, O., Kosunen, I., Ruotsalo, T., Spapé, M.M., Eugster, M.J., Ravaja, N., Kaski, S., Jacucci, G.: Extracting relevance and affect information from physiological text annotation. User Modeling and User-Adapted Interaction **26**(5), 493–520 (2016)
4. Bian, Z., Li, Q., Wang, L., Lu, C., Yin, S., Li, X.: Relative power and coherence of eeg series are related to amnestic mild cognitive impairment in diabetes. Frontiers in aging neuroscience **6**, 11 (2014)
5. Bouaffre, S., Faita-Ainseba, F.: Hemispheric differences in the time-course of semantic priming processes: Evidence from event-related potentials (erps). Brain and Cognition **63**(2), 123–135 (2007)
6. Calbi, M., Siri, F., Heimann, K., Barratt, D., Gallese, V., Kolesnikov, A., Umiltà, M.A.: How context influences the interpretation of facial expressions: a source localization high-density eeg study on the "kuleshov effect". Scientific reports **9**(1), 1–16 (2019)
7. Calhoun, V.: Data-driven approaches for identifying links between brain structure and function in health and disease. Dialogues in clinical neuroscience **20**(2), 87 (2018)
8. Cool, C., Frieder, O., Kantor, P.: Characteristics of text affecting relevance judgments. Proceedings of the 14th National Online Meeting **14** (01 1993)
9. Curran, T.: Brain potentials of recollection and familiarity. Memory & cognition **28**(6), 923–938 (2000)
10. Delorme, A., Makeig, S.: Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. Journal of neuroscience methods **134**(1), 9–21 (2004)
11. Dien, J., Michelson, C.A., Franklin, M.S.: Separating the visual sentence n400 effect from the p400 sequential expectancy effect: Cognitive and neuroanatomical implications. Brain research **1355**, 126–140 (2010)
12. Dimitriadis, S.I., Salis, C., Tarnanas, I., Linden, D.E.: Topological filtering of dynamic functional brain networks unfolds informative chronnectomics: a novel data-driven thresholding scheme based on orthogonal minimal spanning trees (omsts). Frontiers in neuroinformatics **11**, 28 (2017)
13. Elleman, A.M., Oslund, E.L.: Reading comprehension research: Implications for practice and policy. Policy Insights from the Behavioral and Brain Sciences **6**(1), 3–11 (2019)
14. Eugster, M.J., Ruotsalo, T., Spapé, M.M., Barral, O., Ravaja, N., Jacucci, G., Kaski, S.: Natural brain-information interfaces: Recommending information by relevance inferred from human brain signals. Scientific reports **6**, 38580 (2016)

15. Eugster, M.J., Ruotsalo, T., Spapé, M.M., Kosunen, I., Barral, O., Ravaja, N., Jacucci, G., Kaski, S.: Predicting term-relevance from brain signals. In: SIGIR '14. p. 425–434. ACM, NY, USA (2014)
16. Farwell, L.A., Donchin, E.: The truth will out: Interrogative polygraphy ("lie detection") with event-related brain potentials. Psychophysiology **28**(5), 531–547 (1991)
17. Jacucci, G., Barral, O., Daee, P., Wenzel, M., Serim, B., Ruotsalo, T., Pluchino, P., Freeman, J., Gamberini, L., Kaski, S., et al.: Integrating neurophysiologic relevance feedback in intent modeling for information retrieval. JASIST **70**, 917 – 930 (2019)
18. Johnson Jr, R.: The amplitude of the p300 component of the event-related potential: Review and synthesis. Advances in psychophysiology **3**, 69–137 (1988)
19. Johnson Jr, R., Donchin, E.: On how p300 amplitude varies with the utility of the eliciting stimuli. Electroencephalography and Clinical Neurophysiology **44**(4), 424–437 (1978)
20. Kauppi, J.P., Kandemir, M., Saarinen, V.M., Hirvenkari, L., Parkkonen, L., Klami, A., Hari, R., Kaski, S.: Towards brain-activity-controlled information retrieval: Decoding image relevance from meg signals. NeuroImage **112**, 288–298 (2015)
21. Kelly, S.P., O'Connell, R.G.: Internal and external influences on the rate of sensory evidence accumulation in the human brain. Journal of Neuroscience **33**(50), 19434–19441 (2013)
22. Kim, H.H., Kim, Y.H.: Erp/mmr algorithm for classifying topic-relevant and topic-irrelevant visual shots of documentary videos. JASIST **70**(9), 931–941 (2019)
23. Laganaro, M., Perret, C.: Comparing electrophysiological correlates of word production in immediate and delayed naming through the analysis of word age of acquisition effects. Brain Topography **24**(1), 19–29 (2011)
24. LePendu, P., Dou, D., Frishkoff, G.A., Rong, J.: Ontology database: A new method for semantic modeling and an application to brainwave data. In: International Conference on Scientific and Statistical Database Management. pp. 313–330. Springer (2008)
25. Liu, Y., Shi, G., Li, M., Xing, H., Song, Y., Xiao, L., Guan, Y., Han, Z.: Early top-down modulation in visual word form processing: Evidence from an intracranial seeg study. Journal of Neuroscience **41**(28), 6102–6115 (2021)
26. Luck, S.J.: An introduction to the event-related potential technique. MIT press (2014)
27. Meghdadi, A.H., Karić, M., Berka, C.: Eeg analytics: benefits and challenges of data driven eeg biomarkers for neurodegenerative diseases. 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC) pp. 1280–1285 (2019)
28. Mognon, A., Jovicich, J., Bruzzone, L., Buiatti, M.: Adjust: An automatic eeg artifact detector based on the joint use of spatial and temporal features. Psychophysiology **48**(2), 229–240 (2011)
29. Moshfeghi, Y.: Neurasearch: Neuroscience and information retrieval. In: CEUR Workshop Proceedings. vol. 2950, pp. 193–194 (2021)
30. Moshfeghi, Y., Pinto, L.R., Pollick, F.E., Jose, J.M.: Understanding relevance: An fmri study. In: ECIR'13. p. 14–25. Springer-Verlag, Berlin, Heidelberg (2013)
31. Moshfeghi, Y., Pollick, F.E.: Search process as transitions between neural states. In: WWW '18. p. 1683–1692. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE (2018)
32. Moshfeghi, Y., Pollick, F.E.: Neuropsychological model of the realization of information need. JASIST **70**(9), 954–967 (2019)
33. Moshfeghi, Y., Triantafillou, P., Pollick, F.: Towards predicting a realisation of an information need based on brain signals. In: WWW '19. p. 1300–1309. ACM, NY, USA (2019)

34. Moshfeghi, Y., Triantafillou, P., Pollick, F.E.: Understanding information need: An fmri study. In: SIGIR '16. p. 335–344. ACM, NY, USA (2016)
35. O'connell, R.G., Dockree, P.M., Kelly, S.P.: A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. Nature neuroscience **15**(12), 1729 (2012)
36. Paisalnan, S., Moshfeghi, Y., Pollick, F.: Neural correlates of realisation of satisfaction in a successful search process. Proceedings of the Association for Information Science and Technology **58**(1), 282–291 (2021)
37. Paisalnan, S., Pollick, F., Moshfeghi, Y.: Towards understanding neuroscience of realisation of information need in light of relevance and satisfaction judgement. In: International Conference on Machine Learning, Optimization, and Data Science. pp. 41–56. Springer (2021)
38. Pinkosova, Z., McGeown, W.J., Moshfeghi, Y.: The cortical activity of graded relevance. In: SIGIR '20. p. 299–308. ACM, NY, USA (2020)
39. Polich, J.: Updating p300: an integrative theory of p3a and p3b. Clinical neurophysiology **118**(10), 2128–2148 (2007)
40. Ruchkin, D., Sutton, S.: Emitted p300 potentials and temporal unvertainty. Electroencephalography and Clinical Neurophysiology **45**(2), 268–277 (1978)
41. Rutman, A.M., Clapp, W.C., Chadick, J.Z., Gazzaley, A.: Early top–down control of visual processing predicts working memory performance. Journal of cognitive neuroscience **22**(6), 1224–1234 (2010)
42. Saracevic, T.: Relevance: A review of the literature and a framework for thinking on the notion in information science. part iii: Behavior and effects of relevance. JASIST **58**(13), 2126–2144 (2007)
43. Savostyanov, A., Bocharov, A., Astakhova, T., Tamozhnikov, S., Saprygin, A., Knyazev, G.: The behavioral and erp responses to self-and other-referenced adjectives. Brain Sciences **10**(11), 782 (2020)
44. Schmüser, L., Sebastian, A., Mobascher, A., Lieb, K., Tüscher, O., Feige, B.: Data-driven analysis of simultaneous eeg/fmri using an ica approach. Frontiers in neuroscience **8**, 175 (2014)
45. Sitnikova, T., Salisbury, D.F., Kuperberg, G., Holcomb, P.J.: Electrophysiological insights into language processing in schizophrenia. Psychophysiology **39**(6), 851–860 (2002)
46. Sormunen, E.: Liberal relevance criteria of trec-: Counting on negligible documents? In: SIGIR '02. pp. 324–330. ACM (2002)
47. Spironelli, C., Angrilli, A.: Complex time-dependent erp hemispheric asymmetries during word matching in phonological, semantic and orthographical matching judgment tasks. Symmetry **13**(1), 74 (2021)
48. Tagliabue, C.F., Veniero, D., Benwell, C.S., Cecere, R., Savazzi, S., Thut, G.: The eeg signature of sensory evidence accumulation during decision formation closely tracks subjective perceptual experience. Scientific reports **9**(1), 1–12 (2019)
49. Wang, F., He, Y., Pan, J., Xie, Q., Yu, R., Zhang, R., Li, Y.: A novel audiovisual brain-computer interface and its application in awareness detection. Scientific reports **5**(1), 1–12 (2015)
50. Wang, L., Zheng, J., Huang, S., Sun, H.: P300 and decision making under risk and ambiguity. Computational intelligence and neuroscience **2015** (2015)
51. Yang, H., Laforge, G., Stojanoski, B., Nichols, E.S., McRae, K., Köhler, S.: Late positive complex in event-related potentials tracks memory signals when they are decision relevant. Scientific reports **9**(1), 1–15 (2019)
52. Zhitomirsky-Geffet, M., Bar-Ilan, J., Levene, M.: How and why do users change their assessment of search results over time? ASIST **52**(1), 1–4 (2015)