IRIS: Interpretable Retrieval-Augmented Classification for Long Interspersed Document Sequences

Fengnan Li, Elliot D. Hill, Shu Jiang, Jiaxin Gao, Matthew M. Engelhard
Duke University / Durham, NC, USA
{fengnan.li, elliot.d.hill, jiang.shu, jiaxin.gao, m.engelhard}@duke.edu

Abstract

Transformer-based models have achieved state-of-the-art performance in document classification but struggle with long-text processing due to the quadratic computational complexity in the self-attention module. Existing solutions, such as sparse attention, hierarchical models, and key sentence extraction, partially address the issue but still fall short when the input sequence is exceptionally lengthy. To address this challenge, we propose **IRIS** (**I**nterpretable **R**etrieval-Augmented Classification for long **I**nterspersed Document **S**equences), a novel, lightweight framework that utilizes retrieval to efficiently classify long documents while enhancing interpretability. IRIS segments documents into chunks, stores their embeddings in a vector database, and retrieves those most relevant to a given task using learnable query vectors. A linear attention mechanism then aggregates the retrieved embeddings for classification, allowing the model to process arbitrarily long documents without increasing computational cost and remaining trainable on a single GPU. Our experiments across six datasets show that IRIS achieves comparable performance to baseline models on standard benchmarks, and excels in three clinical note disease risk prediction tasks where documents are extremely long and key information is sparse. Furthermore, IRIS provides global interpretability by revealing a clear summary of key risk factors identified by the model. These findings highlight the potential of IRIS as an efficient and interpretable solution for long-document classification, particularly in healthcare applications where both performance and explainability are crucial. [1]

1 Introduction

Transformer-based models (Vaswani, 2017) have achieved state-of-the-art performance in various natural language processing (NLP) tasks, including document classification. At the core of these models is the self-attention mechanism, which allows them to capture dependencies between any tokens in an input sequence. However, this mechanism is computationally expensive, as calculating the full attention matrix scales quadratically in both time and memory with input length. This makes it challenging to apply transformer-based models to long-text processing. For example, BERT (Devlin et al., 2019) and its variants (e.g. Lan, 2019; Liu, 2019; He et al., 2020) are typically pretrained to handle sequences of up to 512 tokens, whereas real-world documents often far exceed this token limit - whether in the form of individual lengthy documents, or collections of documents that need to be processed jointly, e.g., a patient's clinical notes across multiple encounters for disease risk prediction.

Usually, transformer-based models handle long documents in three ways: truncation, sparse attention, and chunking. Truncation, the simplest approach, uses only a fixed portion of the text (e.g., the first 512 tokens), but this risks losing important information. Sparse attention methods (e.g. Beltagy et al., 2020; Zaheer et al., 2020; Guo et al., 2021; Warner et al., 2024) increase the token limit by reducing the computational complexity of the self-attention mechanism, but they still impose an upper bound on the number of tokens. The chunking approach utilizes a hierarchical model (e.g. Pappagari et al., 2019; Lu et al., 2021; Jaiswal and Milios, 2023). These methods split long texts into chunks, process each separately, and then combine the chunk representations. This method circumvents the token limit constraint in theory, but in practice is limited by available memory when dealing with very long inputs.

An important case of long-document classification is disease risk prediction based on patient histories documented in clinical notes. These histo-

[1] Code is available at https://github.com/fengnanli-neo/iris

Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 30263–30283
July 27 - August 1, 2025 ©2025 Association for Computational Linguistics

ries often contain hundreds of thousands of words (Jensen et al., 2017; Rule et al., 2021), far exceeding the capacity of transformer-based models. Additionally, most clinical text is not relevant to a given disease risk prediction task. The small fraction of text containing disease-related information is buried within extensive records, making it inefficient to process all the text indiscriminately. Thus, we need an alternative approach that first selects key information from the histories, then uses it to make predictions.

One potential approach is to first extract key sentences using an unsupervised method. These methods fall into three categories (Papagiannopoulou and Tsoumakas, 2020): statistical (El-Beltagy and Rafea, 2009; Liu et al., 2009; Campos et al., 2020), graph-based (Mihalcea and Tarau, 2004), and embedding-based techniques (Bennani-Smires et al., 2018; Kong et al., 2023). However, in all three cases, the methods are designed to extract sentences to form a succinct summary of the whole document rather than extracting task-specific information for downstream classification. Recent work (Li et al., 2024) has attempted to adapt these methods by filtering out general keyphrases and retaining only those relevant to downstream tasks. However, this approach falls short when critical information is absent from the initially extracted summaries—a common issue when key information for classification is sparse or does not dominate the text. Other strategies (Ding et al., 2020) involve building a separate judge model that is jointly trained with a classification model to evaluate the sentence relevance for classification task. While this approach improves task specificity, it is computationally expensive and impractical for extremely long documents (Park et al., 2022).

Beyond the challenges of handling long input sequences, model interpretability is another crucial concern in long-document classification. Manually reviewing the entire text to understand a model's prediction is often impractical, making it essential to have an efficient method for interpreting model decisions. This is especially important in healthcare applications, where we need models to clearly point to the evidence behind their decisions. Existing interpretability techniques (e.g. Ribeiro et al., 2016; Lundberg, 2017; Sundararajan et al., 2017) for deep learning-based document classification primarily focus on explaining individual predictions by highlighting key words and sentences rather than offering a global summary of key classification factors. For instance, in disease risk prediction, it is not only important to explain why a model made a specific prediction for a single patient but also to provide a clear, comprehensive view of the most critical risk factors across all cases. Additionally, current interpretability methods are often computationally intensive and require post hoc analysis after model training, further increasing complexity.

To address these challenges, we propose **IRIS** (**I**nterpretable **R**etrieval-Augmented Classification for long **I**nterspersed Document **S**equences), a novel, lightweight, interpretable framework for long-document classification. Inspired by the increasing popularity of retrieval-augmented generation (Gao et al., 2023), we adapt this concept for long document classification tasks. Our approach first segments documents into chunks, storing their embeddings in a vector database to enable semantic retrieval. The model architecture is simple and easy to implement: IRIS employs a set of query vectors as the model's learnable parameters, each retrieving k relevant chunk embeddings of a given document from the vector database. Classification is performed solely on these retrieved embeddings, with a linear attention mechanism aggregating information across them. The linear attention mechanism also enables the query vectors to be updated during the training process. Each query vector is progressively refined to specialize in retrieving one key classification factor.

Our model can process arbitrarily long documents by retrieving task-relevant chunks while remaining trainable on a single GPU. Additionally, as suggested by its name, the model's query vectors function like an iris, enabling us to "see" the key factors it prioritizes for classification. In disease risk prediction tasks, this capability is particularly valuable for discovering subtle risk factors that may be scattered throughout lengthy clinical documentation. By inspecting the chunks in our vector database nearest in embedding space to each query, we can directly observe the key factors identified by the model. This approach provides a global summary of key classification factors while eliminating the need for post hoc explanation methods.

Our experimental results show that IRIS performs comparably to commonly used baselines on standard benchmarks and excels in three clinical note disease risk prediction tasks that the other methods cannot solve, where documents are particularly lengthy and key information is sparse—aligning with our model's design. We also highlight the model's favorable interpretability features in the clinical note datasets.

To summarize, the key contributions of our work include:

1. We present a novel framework that adapts retrieval techniques, which are widely used in generative tasks, for document classification.

2. We develop a lightweight, easy-to-implement model whose computational cost in time and memory is invariant to input length, addressing the central challenge of long-document classification.

3. Our model supports both global interpretability, providing a clear summary of key classification factors across all cases, and case-level interpretability, by retrieving relevant chunks from individual documents for specific predictions. This interpretability is achieved during training, eliminating the need for extensive post hoc computations.

4. We evaluate our model across several standard document classification datasets, demonstrating its competitive performance compared to existing approaches.

5. We successfully apply our method to three challenging clinical note disease risk prediction tasks that exceed the capacity of existing methods, showcasing its ability to identify key risk factors from extensive patient clinical note histories while maximizing interpretability.

2 Related Work

To adapt Transformer-based models for long-document processing, various approaches have been explored to extend or circumvent token limits. One straightforward method is sparse attention, initially introduced in Longformer (Beltagy et al., 2020) and BigBird (Zaheer et al., 2020), which combine local and global attention to reduce the quadratic complexity of self-attention computation. These models increase the token limit to 4,096, while the more recent ModernBERT (Warner et al., 2024) adopts an alternative attention mechanism, allowing it to process up to 8,192 tokens.

Other approaches aim to circumvent token limitations altogether. One such method is hierarchical modeling, where a document is split into multiple chunks, each processed separately before aggregating the chunk embeddings. Variants of this method primarily differ in how chunk embeddings are integrated. For example, RoBERT (Pappagari et al., 2019), ToBERT (Pappagari et al., 2019), and ChunkBERT (Jaiswal and Milios, 2023) employ an LSTM, a Transformer encoder, and a TextCNN module, respectively, for chunk aggregation. However, these methods require substantial GPU memory, as processing a large number of chunks becomes increasingly expensive for long documents.

An alternative approach is key sentence extraction, which avoids processing the full document. BERT+TextRank (Park et al., 2022) and Chulo (Li et al., 2024) apply unsupervised keyphrase extraction techniques to select relevant text segments. However, these methods are not optimized for downstream tasks and can fail when task-relevant text is not selected by the extraction algorithm. CogLTX (Ding et al., 2020) extracts text blocks by training a separate model to score text block relevance, filtering out less useful content through an iterative process in which text blocks are removed one by one to observe the resulting impact on model performance. However, this method is highly inefficient, reported to be 104.52 times slower than a standard BERT model when processing documents of approximately 500 words (Park et al., 2022). For significantly longer documents, labeling and ranking the relevance of every text block becomes impractical.

Our method falls under key sentence extraction for long-document classification, but differs from prior work by leveraging retrieval-based techniques, which are widely used in text generation tasks (Lewis et al., 2020). This allows us to efficiently extract task-relevant sentences to circumvent the token limit of transformer-based models.

3 IRIS

3.1 Vector Database

Before training the model, we construct a vector database for semantic retrieval. We begin by splitting all documents into chunks and encoding each chunk into an embedding using a transformer-based model fine-tuned for retrieval. The resulting chunk embeddings are stored in the vector database. Given an input vector, we compute its cosine similarity with chunk embeddings and retrieve the top-k most relevant chunks. In practice, this retrieval process can be accelerated using techniques such as product quantization (Jegou et al., 2010).

3.2 Model Architecture

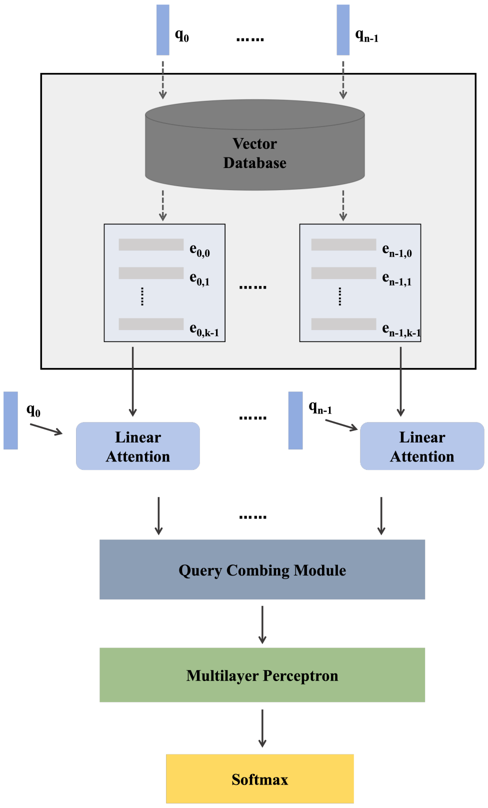

Figure 1: IRIS Model Architecture. The model has n query vectors set as learnable parameters, each responsible for retrieving k chunk embeddings from the vector database. Retrieved embeddings are aggregated through a linear attention mechanism. The dotted arrows in the figure correspond to the retrieval process that is not part of the model's computational graph.

The IRIS model has n query vectors set as learnable parameters, denoted as $q_0, q_1, ..., q_{n-1}$. We explore two initialization strategies for these query vectors: (1) Random Initialization; (2) Cluster-Based Initialization, in which we sample a subset of chunk embeddings in the vector database and apply K-means clustering to partition them into n clusters; the cluster centroids are then used to initialize the query vectors.

Given a single long document (or document sequence) in a batch, each query vector q_i retrieves k embeddings of chunks from that document from the vector database, denoted as $e_{i,0}, e_{i,1}, ..., e_{i,k-1}$. We integrate the retrieved embeddings using a simple linear attention mechanism to generate a single prediction vector v_i, defined as follows:

$$\begin{aligned} v_i &= \sum_{j=0}^{k-1} a_{i,j} \cdot e_{i,j} \\ a_{i,j} &= \frac{\exp(w_{i,j}/T)}{\sum_{j=0}^{k-1} \exp(w_{i,j}/T)} \\ w_{i,j} &= \text{dot}(q_i, e_{i,j}) \end{aligned} \tag{1}$$

We may view this as a special case of an attention mechanism in which the query matrix Q consists of a single query vector q_i, and the key K and value V matrices are the retrieved embeddings $e_{i,0}, ..., e_{i,k-1}$. The softmax temperature T is a hyperparameter controlling attention sharpness. The resulting prediction vector v_i is a weighted sum of the k retrieved chunk embeddings.

In this way, each query vector q_i produces a prediction vector v_i that represents the chunks retrieved by q_i. To obtain a final document representation, we combine $v_0, v_1, ..., v_{n-1}$ into a single prediction vector v^* using one of the following methods: (1) Concatenation; (2) Weighted Sum with a learnable weight corresponding to each v_i; (3) Multihead Attention followed by mean pooling; (4) Transformer Encoder layer (Vaswani, 2017) followed by mean pooling. The choice of combination method is treated as a tunable hyperparameter. The a multi-layer perceptron (MLP) prediction head is then applied to v^*.

The model structure of IRIS is quite simple. The core is the linear attention mechanism in Equation 1, which enables query vectors to be refined dynamically during training. Initially, each query vector retrieves arbitrary chunk embeddings, some of which may be informative for classification, while others are not. During each training step, the model learns to assign higher weights to the most predictive of the k embeddings it retrieves, ensuring information relevant to the prediction task is retained in the prediction vector v_i.

To achieve this, q_i is updated to increase its similarity (dot product) with the useful embeddings, moving it closer to key classification factors and allowing it to retrieve a more relevant set of chunks in the next training step. As training progresses, q_i gradually converges to retrieve the most relevant chunk embeddings within a particular region of embedding space, refining its selection of task-critical information. The exact gradient update rule for q_i is provided in Equation 2.

The model architecture is illustrated in Figure 1.

3.3 Hyperparameters for Improving Interpretability

One of the key features of IRIS is its global interpretability. Ideally, each query vector should be responsible for retrieving a specific type of risk factor, providing a clear overview of the key factors identified. However, in practice, two common challenges hinder model interpretability: (1) a single query vector may retrieve a mix of multiple factors instead of focusing on one, making it difficult to obtain a comprehensive overview of all retrieved factors; (2) different query vectors may converge toward similar directions, resulting in redundant retrieval of the same content and limiting the diversity of key factors detected. In this section, we present and discuss our approach to mitigate these issues.

First, to enhance query specificity, we can reduce T, the softmax temperature in Equation 1. The gradient of query vector q_i is given by:

$$\nabla_{q_i}\mathcal{L} = \frac{1}{T}\left[\sum_{j=0}^{k-1} a_{i,j}\left(\frac{\partial \mathcal{L}}{\partial v_i}\cdot e_{i,j}\right)\cdot e_{i,j} - \left(\frac{\partial \mathcal{L}}{\partial v_i}\cdot v_i\right)\cdot v_i\right] \tag{2}$$

where $\mathcal{L}$ is the loss function. See Appendix A for a detailed derivation of this query vector gradient.

By selecting a small temperature T, we create a sharper distribution of attention scores $a_{i,j}$, increasing the relative contribution of embeddings closest to q_i to its gradient, and reducing the contribution of retrieved embeddings that are farther away. Intuitively, this makes q_i more specific by concentrating its focus on smaller set of embeddings.

Second, to prevent different query vectors from retrieving redundant content, we add a loss term that penalizes high pairwise cosine similarities (dot products) among normalized query vectors:

$$\mathcal{L}_{\text{penalty}} = \lambda \sum_{i=0}^{k-1}\sum_{j=0}^{k-1} \textbf{ReLU}(\text{dot}(q_i^*, q_j^*) - thres) \tag{3}$$

where q_i^* and q_j^* are the L2-normalized query vectors. The hyperparameter λ controls the magnitude of the penalty, and $thres$ sets the similarity threshold. The goal is not to minimize all pairwise similarities but to ensure that query vectors remain sufficiently distinct to retrieve diverse content. By applying a ReLU function, we penalize only similarities exceeding the threshold, while lower similarities do not contribute to the penalty term.

3.4 Two Stage Model for Fine-tuning

The previous sections describe the IRIS model, which utilizes fixed embeddings stored in a vector database. However, for relatively short documents or cases where sufficient GPU memory is available to store the computational graph of multiple embeddings, a two-stage model can be used to fine-tune the chunk embeddings.

The two-stage model retains the same overall structure as the IRIS model, with the only difference being that it incorporates the embedding encoder within the model architecture, allowing chunk embeddings to be dynamically updated during training. During each training iteration, we generate new chunk embeddings for all chunk text in the batch using the model's encoder. These chunk embeddings are detached from the computational graph and used to update the vector database. Meanwhile, the chunk embeddings with gradients are stored in a dictionary. The retrieval process retrieves the indices of the corresponding chunks. These indices serve as dictionary keys to obtain the retrieved chunk embeddings along with their gradient information. The retrieved embeddings are then used in the same classification process as previously discussed. This two-stage approach enables fine-tuning of the encoder, potentially enhancing model performance.

4 Experimental Setup

4.1 Datasets

The IRIS model is designed for classifying extremely long documents where key information is sparse. In this paper, we evaluate our model using three standard long document classification datasets and three medical datasets featuring extremely long clinical notes for disease risk prediction, which better align with our model's intended application.

IMDb (Maas et al., 2011) A collection of movie reviews for binary sentiment analysis. The dataset contains 50,000 documents, with an average length of 389 words per document.

Hyperpartisan (Kiesel et al., 2019) A binary classification dataset consisting of 1,273 news articles labeled as hyperpartisan or non-hyperpartisan. The average document length is 586 words.

arXiv (**He et al., 2019**) A multi-class dataset for classifying research papers by subject area based on their full text. In our experiments, we select three subjects: cs.AI (Artificial Intelligence), cs.DS (Data Structures), and cs.PL (Programming Languages). The dataset contains 10,032 articles, with an average document length of 6,953 words.

MIMIC-IV (**Johnson et al., 2023**) A publicly accessible electronic health record dataset from which we constructed a binary classification task to predict essential (primary) hypertension (ICD-9 Code 4019, ICD-10 Code I-10) using de-identified discharge notes. We concatenated all discharge notes per patient into a single long document, excluded patients with fewer than 20,000 words of clinical text to ensure sufficient document length for our model evaluation, and randomly sampled 500 cases and 500 control patients for a balanced dataset.

ADHD An institutionally controlled binary classification dataset for early ADHD (Attention Deficit Hyperactivity Disorder) risk prediction using clinical notes from Electronic Health Records at Duke University Hospital. Our goal is to predict ADHD risk at an early age, prior to any formal diagnosis. To this end, we include all clinical notes recorded for each patient before the age of two, when ADHD diagnoses have typically not yet been made (Loh et al., 2025). The study cohort consists of 3,990 patients born between 2014 and 2022, 620 of whom were later diagnosed with ADHD, while the remaining patients were censored. The average document length per patient is approximately 37,000 words.

Autism An institutionally controlled binary classification dataset for early age autism risk prediction using clinical notes from Electronic Health Records at Duke University Hospital. The dataset includes 4,987 patients born between 2014 and 2022, of whom 751 have been diagnosed with autism. To prevent label leakage and align with the goal of early prediction, we only retain clinical notes recorded before age 1.5 years—a critical period when most autism diagnoses have not yet been made (Loh et al., 2025). The average total document length per patient (concatenating all notes) is 39,423 words.

All datasets are split into training, validation, and test sets by an 8:1:1 ratio using a random seed of 42.

4.2 Models

In addition to the two proposed models (IRIS with fixed embeddings and two-stage IRIS), we implement five baseline models for comparison: **BERT** (Devlin et al., 2019) – standard BERT base model; **BERT + TextRank** (Park et al., 2022) – augments the first 512 tokens with up to 512 additional tokens selected using TextRank; **BERT + Random** (Park et al., 2022) – augments the first 512 tokens with up to 512 tokens from randomly selected sentences; **Longformer** (Beltagy et al., 2020) – the transformer-based model that employs sparse attention with a token limit of 4,096; **ModernBERT** (Warner et al., 2024) - the recently released encoder-only model with a maximum sequence length of 8,192 tokens; **ToBERT** (Pappagari et al., 2019) – a hierarchical model that processes multiple document chunks using BERT and employs a transformer encoder to combine the chunk embeddings.

All models are evaluated on the Hyperpartisan and IMDb datasets. Due to GPU memory constraints, ToBERT and two-stage IRIS are not tested on the four long document datasets (arXiv, MIMIC-IV, ADHD, and Autism). For the three medical datasets that contain exceptionally long clinical notes, BERT-based baselines (BERT, BERT + TextRank, BERT + Random) are unsuitable due to the limited number of tokens they can process. Instead, we only run Longformer and ModernBERT, the baseline models with the largest token limits, on these datasets for comparison. We train all models for 20 epochs, selecting the best-performing model based on validation set performance and reporting its performance on the test dataset. In all cases, validation performance had plateaued or begun to decline before 20 epochs.

For the standard benchmarks (IMDb, Hyperpartisan, and arXiv), BERT-based models process the first 512 tokens, while Longformer and ModernBERT process the first 4,096 and 8,192 tokens, respectively. For the three medical datasets (MIMIC-IV, ADHD, and Autism), we evaluated two truncation strategies for both ModernBERT and Longformer—taking either the first or last 4,096/8,192 tokens—and report the better-performing option for each dataset.

4.3 Vector Database

For the IMDb, Hyperpartisan, arXiv, and MIMIC-IV datasets, each document is split into chunks of 80 words with a 16-word overlap between consecutive chunks. We use the sentence transformer `all-mpnet-base-v2` (MPNET)[2] as the encoder to generate chunk embeddings and store them into a vector database. For the ADHD and Autism datasets, documents are split into chunks of 100 words with a 20-word overlap, and `e5-mistral-7b-instruct`[3] (Wang et al., 2024) is used as the encoder.

5 Results and Analysis

5.1 Model Performance

The model performance across all datasets is summarized in Table 1. We report the mean performance metrics over 5 runs. Although our model is primarily designed for extremely long document classification, it achieves performance comparable to the baseline models on the IMDb and Hyperpartisan datasets, which contain relatively short documents. Additionally, the two-stage IRIS model demonstrates improved performance over the fixed embedding version.

In the arXiv dataset, despite its significantly longer documents, each article begins with an abstract that serves as a high-quality summary of the entire paper. As a result, most of the key information for subject classification is retained within the first 512 tokens, allowing BERT-based models to perform well.

What impose a real challenge to transformer-based baselines are the medical datasets (MIMIC-IV, ADHD, Autism), where documents are extremely long, yet task-relevant information is sparse. Most baseline models cannot be applied to this classification task, and even Longformer and ModernBERT, which process the most tokens among the baselines, fall far short compared to IRIS. In addition to achieving a substantial performance advantage, the IRIS model is also far more efficient due to its lightweight structure.

5.2 Model Interpretability

Model interpretability is essential in clinical practice, where trust and transparency are highly valued. Here, we take the Autism dataset as an example to illustrate the interpretability features of IRIS. For a trained model, we extract its query vectors and use them to perform retrieval on the vector database to examine the types of chunks each query vector learns to extract; which represent the risk factors identified by the model.

A one-word description of the risk factor retrieved by each query vector, based on our interpretation of the retrieved chunks, is provided in Table 2. The one-word descriptions are summarization from Appendix B.1, which includes the 10 closest clinical note chunks retrieved by each query vector. Additionally, we present query vector importance scores in Table 2, obtained by interpolating each query vector and aggregating their contributions across the dataset using the integrated gradients method (Sundararajan et al., 2017).

Most of the risk factors identified by the model align with expectations. For example, skill development (query 1) and late talking (query 7) are known early signs of autism. Other factors like low birthweight (query 0) are also known early correlates of autism (Lampi et al., 2012). Less is known about the relationship between autism and ophthalmological findings such as esotropia, exotropia, amblyopia, and myopia (query 5), but some previous literature has suggested an association (e.g. Kaplan et al., 1999; Milne et al., 2009). This example demonstrates IRIS's potential for uncovering descriptive risk factors not documented in structured data fields and thus overlooked by traditional approaches. However, some extracted factors are unexpected or controversial, such as dental varnish (query 1) and immunization (query 2). It is important to note that the factors identified by the model's query vectors do not imply causal relationships, and may instead reflect the influence of confounding variables. For instance, dental varnish is typically applied only in children who do not have access to (flouridated) municipal water, suggesting that this finding may serve here as a proxy for rurality. Similarly, immunization is an effective proxy for parental engagement with the health system, which is linked to higher diagnosis rates. While our model has the potential to uncover previously unrecognized risk factors, careful interpretation is required, and clinical validation remains essential.

Similar interpretable features are observed across the other two medical datasets. For the ADHD dataset, the risk factors extracted by IRIS's query vectors include developmental screening results, sleep concerns, BPSC (Baby Pediatric Symptom Checklist) scores, maternal mental health indicators, and recurrent fevers/infections. For the

[2] https://huggingface.co/sentence-transformers/all-mpnet-base-v2

[3] https://huggingface.co/intfloat/e5-mistral-7b-instruct

Model	Accuracy (%)				AUC (%)	
	IMDb	Hyper-Partisan	arXiv	MIMIC-IV	ADHD	Autism
BERT	92.56	85.94	90.90	-	-	-
BERT+TextRank	93.12	82.81	94.68	-	-	-
BERT+Random	92.88	82.97	**95.80**	-	-	-
Longformer	94.88	84.38	94.39	53.00	77.54	70.89
ModernBERT	**95.66**	86.60	94.25	67.40	76.50	72.13
ToBERT	86.90	79.61	-	-	-	-
IRIS (fixed-embed)	91.54	**86.72**	95.66	**75.20**	**85.64**	**82.35**
IRIS (two-stage)	93.14	**86.72**	-	-	-	-

Table 1: Test performance across datasets. Accuracy is reported for IMDb, Hyperpartisan, arXiv, and MIMIC-IV datasets, while AUC is used for the highly imbalanced ADHD and Autism datasets. Bold values indicate the best performance on each dataset.

MIMIC-IV dataset, retrieved clinical note chunks are primarily related to medications and treatments associated with hypertension management.

On standard benchmarks, we also observe interpretable patterns. For instance, when training a model with two query vectors on the Hyperpartisan dataset, we observe that one query vector predominantly retrieves article chunks related to scandals involving Hillary and Bill Clinton, while the other focuses on criticism of Trump's presidency, along with discussions on immigration policies and LGBT issues associated with his administration (see Appendix B.2 for details). However, the model's interpretability is notably less stable on non-medical datasets compared to the clinical note datasets. In many cases, some query vectors retrieve chunks without clear, discernible patterns, making it difficult to determine the specific content they have learned to prioritize.

One possible explanation is that the smallest unit influencing query vector refinement is a chunk embedding. If a chunk contains mixed information without a clear focus — a common occurrence in many datasets — it becomes challenging for query vectors to specialize in retrieving distinct content types. In contrast, clinical note datasets often follow structured templates, with each chunk typically centering on a single topic. This structured nature makes it easier for query vectors to converge toward retrieving specific risk factors.

5.3 Hyperparameter Effects

The strategies for improving interpretability presented in Section 3.3 proved effective. When using a softmax temperature of $T = 1$ and omitting the query similarity penalty term in the loss function, we often observe that different query vectors retrieve similar content, or that a single query vector retrieves a mix of risk factors. For example, in the Autism dataset, the top five chunks retrieved by one vector may include notes on both developmental concerns and ophthalmological findings. However, these issues are avoided when setting $\lambda = 0.1$, $thres = 0.4$, and $T = 0.1$, ensuring that each query vector retrieves a distinct and specific risk factor, as shown in Table 2 and Appendix B.1. Additionally, we find that these hyperparameters have little impact on overall model performance, but significantly improve interpretability.

The query initialization methods discussed in Section 3.2 have minimal impact on both model performance and interpretability. Simply initializing query vectors with random values is sufficient to achieve good results.

6 Ablation Study

The key component of IRIS is the set of learnable query vectors, which are optimized during training to retrieve task-relevant chunks. In the ablation study, we replace these with fixed random vectors, thereby removing the model's ability to adapt its retrieval strategy. This modification forces the model

	Query 0	Query 1	Query 2	Query 3
Retrieved Factor	Birth Weight	Dental Varnish	Skill Development	Immunization
Query Importance	0.09	0.06	0.11	0.16
	Query 4	Query 5	Query 6	Query 7
Retrieved Factor	Referral to Specialty Care	Ophthalmology	Provider Interactions	Late Talking
Query Importance	0.08	0.06	0.25	0.19

Table 2: "Risk factors" identified by each query vector on the Autism Dataset and the query vector importance scores. Query importance scores are obtained by integrated gradients.

to rely on randomly extracted chunks. The comparison reveals the importance of task-specific retrieval, both in terms of predictive performance and model interpretability.

In the Autism dataset, we use 8 query vectors and vary k, the number of chunk embeddings retrieved by each query vector. The experimental results are presented in Table 3. Results show that IRIS with learned query vectors consistently achieves higher AUC compared to the fixed-vector variant, especially when k is small. Importantly, the model interpretability feature is also lost when using random query vectors. In contrast, IRIS's learned queries allow us to directly inspect and identify meaningful risk factors from retrieved chunks.

k	Fixed Query	Learned Query
2	67.1	**76.3**
4	71.7	**78.7**
8	76.5	**82.5**
16	79.8	**81.9**
32	79.0	**82.6**

Table 3: Comparison of AUC (%) for fixed versus learned query vectors on the Autism dataset. The better results are in bold.

7 Conclusion

In this work, we introduced IRIS, a novel, interpretable framework for retrieval-augmented long-document classification. By leveraging a vector database for semantic retrieval and learning task-specific query vectors during training, IRIS efficiently identifies and prioritizes relevant information while maintaining a lightweight architecture. Our experiments across six datasets demonstrate that IRIS achieves competitive performance compared to strong baselines on standard benchmarks, and it significantly outperforms existing methods on three clinical risk prediction tasks, where documents (clinical notes) are extremely long and key information is sparse. Notably, IRIS provides global interpretability by revealing the specific risk factors and clinical patterns the model prioritizes for classification, eliminating the need for post hoc explanation methods. This capability is particularly valuable in healthcare applications, where interpretability is important to promote trust among clinical users, and identifying specific descriptive findings that predict a given condition can provide clinically relevant insight. Future work will explore enhancing query vector specialization and expanding the applicability of IRIS to other domains requiring interpretable long-text processing.

Acknowledgments

This work was supported the National Institute of Mental Health (K01MH127309; PI Matthew Engelhard). We also thank Bingkun Wang for his contribution during the rebuttal period and anonymous ARR reviewers for useful feedback.

Limitations

Our model is designed specifically for long-document classification tasks where relevant information is sparse, such as disease risk prediction from a large collection of clinical notes. When applied to shorter documents, IRIS maintains competitive performance and interpretability, but benefits over alternative methods are diminished.

Additionally, IRIS is not well-suited for multi-class or multi-label classification with a large number of classes. Such tasks require a high number of query vectors to capture key classification factors for each class, particularly if we aim to maintain interpretability by assigning each query vector to a distinct factor. While our model is designed for efficiency, an excessive number of query vectors can introduce computational bottlenecks, as the retrieval process becomes increasingly costly when handling numerous retrieval operations.

References

Iz Beltagy, Matthew E Peters, and Arman Cohan. 2020. Longformer: The long-document transformer. *arXiv preprint arXiv:2004.05150*.

Kamil Bennani-Smires, Claudiu Musat, Andreea Hossmann, Michael Baeriswyl, and Martin Jaggi. 2018. Simple unsupervised keyphrase extraction using sentence embeddings. In *Proceedings of the 22nd Conference on Computational Natural Language Learning*, pages 221–229, Brussels, Belgium. Association for Computational Linguistics.

Ricardo Campos, Vítor Mangaravite, Arian Pasquali, Alípio Jorge, Célia Nunes, and Adam Jatowt. 2020. Yake! keyword extraction from single documents using multiple local features. *Information Sciences*, 509:257–289.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Ming Ding, Chang Zhou, Hongxia Yang, and Jie Tang. 2020. Cogltx: Applying bert to long texts. *Advances in Neural Information Processing Systems*, 33:12792–12804.

Samhaa R El-Beltagy and Ahmed Rafea. 2009. Kpminer: A keyphrase extraction system for english and arabic documents. *Information systems*, 34(1):132–144.

Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, and Haofen Wang. 2023. Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*.

Mandy Guo, Joshua Ainslie, David Uthus, Santiago Ontanon, Jianmo Ni, Yun-Hsuan Sung, and Yinfei Yang. 2021. Longt5: Efficient text-to-text transformer for long sequences. *arXiv preprint arXiv:2112.07916*.

Jun He, Liqun Wang, Liu Liu, Jiao Feng, and Hao Wu. 2019. Long document classification from local word glimpses via recurrent attention learning. *IEEE Access*, 7:40707–40718.

Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. 2020. Deberta: Decoding-enhanced bert with disentangled attention. *arXiv preprint arXiv:2006.03654*.

Aman Jaiswal and Evangelos Milios. 2023. Breaking the token barrier: Chunking and convolution for efficient long text classification with bert. *arXiv preprint arXiv:2310.20558*.

Herve Jegou, Matthijs Douze, and Cordelia Schmid. 2010. Product quantization for nearest neighbor search. *IEEE transactions on pattern analysis and machine intelligence*, 33(1):117–128.

Kasper Jensen, Cristina Soguero-Ruiz, Karl Oyvind Mikalsen, Rolv-Ole Lindsetmo, Irene Kouskoumvekaki, Mark Girolami, Stein Olav Skrovseth, and Knut Magne Augestad. 2017. Analysis of free text in electronic health records for identification of cancer patient trajectories. *Scientific reports*, 7(1):46226.

Alistair EW Johnson, Lucas Bulgarelli, Lu Shen, Alvin Gayles, Ayad Shammout, Steven Horng, Tom J Pollard, Sicheng Hao, Benjamin Moody, Brian Gow, et al. 2023. Mimic-iv, a freely accessible electronic health record dataset. *Scientific data*, 10(1):1.

Melvin Kaplan, Bernard Rimland, and Stephen M Edelson. 1999. Strabismus in autism spectrum disorder. *Focus on Autism and Other Developmental Disabilities*, 14(2):101–105.

Johannes Kiesel, Maria Mestre, Rishabh Shukla, Emmanuel Vincent, Payam Adineh, David Corney, Benno Stein, and Martin Potthast. 2019. SemEval-2019 task 4: Hyperpartisan news detection. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 829–839, Minneapolis, Minnesota, USA. Association for Computational Linguistics.

Aobo Kong, Shiwan Zhao, Hao Chen, Qicheng Li, Yong Qin, Ruiqi Sun, and Xiaoyan Bai. 2023. PromptRank: Unsupervised keyphrase extraction using prompt. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 9788–9801, Toronto, Canada. Association for Computational Linguistics.

Katja M Lampi, Liisa Lehtonen, Phuong Lien Tran, Auli Suominen, Venla Lehti, P Nina Banerjee, Mika Gissler, Alan S Brown, and Andre Sourander. 2012. Risk of autism spectrum disorders in low birth weight and small for gestational age infants. *The Journal of pediatrics*, 161(5):830–836.

Zhenzhong Lan. 2019. Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942*.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474.

Yan Li, Soyeon Caren Han, Yue Dai, and Feiqi Cao. 2024. Chulo: Chunk-level key information representation for long document processing. *arXiv preprint arXiv:2410.11119*.

Yinhan Liu. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 364.

Zhiyuan Liu, Peng Li, Yabin Zheng, and Maosong Sun. 2009. Clustering to find exemplar terms for keyphrase extraction. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 257–266, Singapore. Association for Computational Linguistics.

De Rong Loh, Elliot D Hill, Nan Liu, Geraldine Dawson, and Matthew M Engelhard. 2025. Limitations of binary classification for long-horizon diagnosis prediction and advantages of a discrete-time time-to-event approach: Empirical analysis. *JMIR AI*, 4:e62985.

Jinghui Lu, Maeve Henchion, Ivan Bacher, and Brian Mac Namee. 2021. A sentence-level hierarchical bert model for document classification with limited labelled data. In *Discovery Science: 24th International Conference, DS 2021, Halifax, NS, Canada, October 11–13, 2021, Proceedings 24*, pages 231–241. Springer.

Scott Lundberg. 2017. A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874*.

Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. 2011. Learning word vectors for sentiment analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 142–150, Portland, Oregon, USA. Association for Computational Linguistics.

Rada Mihalcea and Paul Tarau. 2004. TextRank: Bringing order into text. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pages 404–411, Barcelona, Spain. Association for Computational Linguistics.

Elizabeth Milne, Helen Griffiths, David Buckley, and Alison Scope. 2009. Vision in children and adolescents with autistic spectrum disorder: Evidence for reduced convergence. *Journal of autism and developmental disorders*, 39(7):965–975.

Eirini Papagiannopoulou and Grigorios Tsoumakas. 2020. A review of keyphrase extraction. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(2):e1339.

Raghavendra Pappagari, Piotr Zelasko, Jesús Villalba, Yishay Carmiel, and Najim Dehak. 2019. Hierarchical transformers for long document classification. In *2019 IEEE automatic speech recognition and understanding workshop (ASRU)*, pages 838–844. IEEE.

Hyunji Park, Yogarshi Vyas, and Kashif Shah. 2022. Efficient classification of long documents using transformers. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 702–709, Dublin, Ireland. Association for Computational Linguistics.

Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144.

Adam Rule, Steven Bedrick, Michael F Chiang, and Michelle R Hribar. 2021. Length and redundancy of outpatient progress notes across a decade at an academic medical center. *JAMA Network Open*, 4(7):e2115334–e2115334.

Mukund Sundararajan, Ankur Taly, and Qiqi Yan. 2017. Axiomatic attribution for deep networks. In *International conference on machine learning*, pages 3319–3328. PMLR.

A Vaswani. 2017. Attention is all you need. *Advances in Neural Information Processing Systems*.

Liang Wang, Nan Yang, Xiaolong Huang, Linjun Yang, Rangan Majumder, and Furu Wei. 2024. Improving text embeddings with large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11897–11916, Bangkok, Thailand. Association for Computational Linguistics.

Benjamin Warner, Antoine Chaffin, Benjamin Clavié, Orion Weller, Oskar Hallström, Said Taghadouini, Alexis Gallagher, Raja Biswas, Faisal Ladhak, Tom Aarsen, et al. 2024. Smarter, better, faster, longer: A modern bidirectional encoder for fast, memory efficient, and long context finetuning and inference. *arXiv preprint arXiv:2412.13663*.

Manzil Zaheer, Guru Guruganesh, Kumar Avinava Dubey, Joshua Ainslie, Chris Alberti, Santiago Ontanon, Philip Pham, Anirudh Ravula, Qifan Wang, Li Yang, et al. 2020. Big bird: Transformers for longer sequences. *Advances in neural information processing systems*, 33:17283–17297.

A Query Vector Gradients Analysis

This section provides the derivation for Equation 2

Definitions:

$$\begin{aligned}
z_i &= \frac{1}{T}\,\big(q^\top e_i\big),\\
\alpha_i &= \frac{\exp(z_i)}{\sum_{m=0}^{k-1}\exp(z_m)} \;=\; \mathrm{softmax}\Big(\tfrac{q^\top e_0}{T},\ldots,\tfrac{q^\top e_{k-1}}{T}\Big)_i,\\
v &= \sum_{i=0}^{k-1}\alpha_i\,e_i,\\
\mathcal{L} &= \mathcal{L}(v).
\end{aligned}$$

Chain Rule:

$$\begin{aligned}
\nabla_q\mathcal{L} = \nabla_v\mathcal{L}\,\frac{\partial v}{\partial q} \;&=\; \nabla_v\mathcal{L}\;\cdot\;\Big(\sum_{i=0}^{k-1}\tfrac{\partial\alpha_i}{\partial q}\,e_i\Big)\\
&= \sum_{i=0}^{k-1}\Big(\nabla_v\mathcal{L}\;\cdot\;\tfrac{\partial\alpha_i}{\partial q}\Big)\,e_i.
\end{aligned}$$

Derivative of α_i w.r.t. q:

$$\begin{aligned}
&\alpha_i = \mathrm{softmax}\Big(\tfrac{q^\top e_0}{T},\ldots,\tfrac{q^\top e_{k-1}}{T}\Big)_i\\
&\quad\Longrightarrow\quad \frac{\partial\alpha_i}{\partial q} \;=\; \frac{\alpha_i}{T}\Big(e_i \;-\; \sum_{j=0}^{k-1}\alpha_j\,e_j\Big) \;=\; \frac{\alpha_i}{T}\,\big(e_i - v\big),
\end{aligned}$$

Combine the Results:

$$\begin{aligned}
\nabla_q\mathcal{L} &= \sum_{i=0}^{k-1}\Big(\nabla_v\mathcal{L}\;\cdot\;\tfrac{\alpha_i}{T}(e_i - v)\Big)\,e_i\\
&= \frac{1}{T}\sum_{i=0}^{k-1}\alpha_i\Big[(\nabla_v\mathcal{L}\cdot e_i) \;-\; (\nabla_v\mathcal{L}\cdot v)\Big]\,e_i\\
&= \frac{1}{T}\Big[\sum_{i=0}^{k-1}\alpha_i\,\big(\nabla_v\mathcal{L}\cdot e_i\big)\,e_i \;-\; \big(\nabla_v\mathcal{L}\cdot v\big)\Big(\sum_{i=0}^{k-1}\alpha_i\,e_i\Big)\Big]\\
&= \frac{1}{T}\Big[\sum_{i=0}^{k-1}\alpha_i\,\big(\nabla_v\mathcal{L}\cdot e_i\big)\,e_i \;-\; \big(\nabla_v\mathcal{L}\cdot v\big)\,v\Big].
\end{aligned}$$

Final Equation:

$$\nabla_q\mathcal{L} \;=\; \frac{1}{T}\Big[\sum_{i=0}^{k-1}\alpha_i\,\big(\nabla_v\mathcal{L}\cdot e_i\big)\,e_i \;-\; \big(\nabla_v\mathcal{L}\cdot v\big)\,v\Big].$$

B Retrieved Chunk Text

B.1 Autism Dataset

We present the top 10 clinical note chunks retrieved for each of the model's query vectors. Note text has been de-identified and reviewed before being exported from a protected environment. Some retrieved chunks are identical because they originate from clinical notes that follow the same template.

Query 0:

1. Birthweight: 3.61 kg (7 lb 15.3 oz) Length: 50.1 cm (19.72") Head Circumference: 33.1 cm (13.03") Today's weight:3.504 kg (7 lb 11.6 oz) Weight change from birth: -3% Weight: 62 %ile (Z= 0.30) based on WHO (Girls, 0-2 years) weight-for-age data using vitals from [DATE]. Length: 58 %ile (Z= 0.19) based on WHO (Girls, 0-2 years) Length-for-age data based on Length recorded on [DATE]. HC: 17 %ile (Z= -0.96) based on WHO (Girls, 0-2 years) head circumference-for-age based on Head Circumference recorded on [DATE]. Weight trend: Wt Readings from Last 3 Encounters: [DATE] 3.504 kg (7 lb 11.6 oz) (62

2. Birthweight: 3.12 kg (6 lb 14.1 oz) Length: 48.5 cm (19.09") Head Circumference: 34.6 cm (13.62") Today's weight:3 kg (6 lb 9.8 oz) Weight change from birth: -4% Weight: 17 %ile (Z= -0.96) based on WHO (Boys, 0-2 years) weight-for-age data using vitals from [DATE]. Length: 16 %ile (Z= -0.98) based on WHO (Boys, 0-2 years) Length-for-age data based on Length recorded on [DATE]. HC: 46 %ile (Z= -0.11) based on WHO (Boys, 0-2 years) head circumference-for-age based on Head Circumference recorded on [DATE]. Weight trend: Wt Readings from Last 3 Encounters: [DATE] 3 kg (6 lb 9.8 oz) (17

3. Change from Birthweight: -8% Birth Weight: 3.32 kg (7 lb 5.1 oz) ([DATE]) 4 %ile (Z= -1.80) based on WHO (Boys, 0-2 years) weight-for-age data using vitals from [DATE]. Classification: AGA Birth Length: 54 cm (21.26") ([DATE]) 98 %ile (Z= 2.10) based on WHO (Boys, 0-2 years) length-for-age data using vitals from [DATE]. Birth HC: 36 cm (14.17") 87 %ile (Z= 1.15) based on WHO (Boys, 0-2 years) head circumference-for-age data using vitals from [DATE]. Current Weight: 3.14 kg (6 lb 14.8 oz) ([DATE]) 4 %ile (Z= -1.80) based on WHO (Boys, 0-2 years) weight-for-age data using

4. kg/m² Birthweight: 2.87 kg (6 lb 5.2 oz) Length: 46 cm (18.11") Head Circumference: 34 cm (13.39") Today's weight:2.93 kg (6 lb 7.4 oz) Weight change from birth: 2% (NEWT) Weight: 17 %ile (Z= -0.94) based on WHO (Girls, 0-2 years) weight-for-age

data using vitals from [DATE]. Length: 2 %ile (Z= -2.00) based on WHO (Girls, 0-2 years) Length-for-age data based on Length recorded on [DATE]. HC: 42 %ile (Z= -0.19) based on WHO (Girls, 0-2 years) head circumference-for-age based on Head Circumference recorded on [DATE]. Weight trend: Wt Readings from Last 3 Encounters: [DATE] 2.93 kg (6 lb 7.4

5. 13.29 kg/m^2 Birthweight: 2.44 kg (5 lb 6.1 oz) Length: 42 cm (16.54") Head Circumference: 32.5 cm (12.8") Today's weight:(!) 2.345 kg (5 lb 2.7 oz) Weight change from birth: -4% Weight: <1%ile (Z= -2.39) based on WHO (Girls, 0-2 years) weight-for-age data using vitals from [DATE]. Length: <1%ile (Z= -4.13) based on WHO (Girls, 0-2 years) Length-for-age data based on Length recorded on [DATE]. HC: 7%ile (Z= -1.46) based on WHO (Girls, 0-2 years) head circumference-for-age based on Head Circumference recorded on [DATE]. Weight trend: Wt Readings from Last 3 Encounters: [DATE] (!) 2.345 kg (5 lb 2.7 oz).

6. (plotted on WHO Growth Chart) Birth Weight: 3.34 kg (7 lb 5.8 oz) 49th percentile wt-for-age Classification: AGA Change from Birthweight: -1% Current Weight: 3.29 kg (7 lb 4.1 oz) ([DATE]) 43%ile (Z= -0.18) based on WHO (Boys, 0-2 years) weight-for-age data using vitals from [DATE]. Current Length: 55 cm (21.65") ([DATE]) >99%ile (Z= 2.70) based on WHO (Boys, 0-2 years) length-for-age data using vitals from [DATE]. Current HC: 31 cm (12.21") <1%ile (Z= -2.72) based on WHO (Boys, 0-2 years) head circumference-for-age data using vitals from [DATE]. Wt/length: <1%ile (Z= -3.79) based on WHO growth standards.

7. kg/m^2 Birthweight: 3.37 kg (7 lb 6.9 oz) Length: 50 cm (19.69") Head Circumference: 35 cm (13.78") Toda's weight:3.189 kg (7 lb 0.5 oz) Weight change from birth: -5% (NEWT) Weight: 31 %ile (Z= -0.49) based on WHO (Girls, 0-2 years) weight-for-age data using vitals from [DATE]. Length: 49 %ile (Z= -0.02) based on WHO (Girls, 0-2 years) Length-for-age data based on Length recorded on [DATE]. HC: 69 %ile (Z= 0.50) based on WHO (Girls, 0-2 years) head circumference-for-age based on Head Circumference recorded on [DATE]. Weight trend: Wt Readings from Last 3 Encounters: [DATE] 3.189 kg (7 lb 0.5

8. Change from Birthweight: 13% Birth Weight: 3.32 kg (7 lb 5.1 oz) ([DATE]) 4 %ile (Z= -1.70) based on WHO (Boys, 0-2 years) weight-for-age data using vitals from [DATE]. Classification: AGA Birth Length: 54 cm (21.26") ([DATE]) 98 %ile (Z= 2.10) based on WHO (Boys, 0-2 years) length-for-age data using vitals from [DATE]. Birth HC: 36 cm (14.17") 87 %ile (Z= 1.15) based on WHO (Boys, 0-2 years) head circumference-for-age data using vitals from [DATE]. Current Weight: 3.85 kg (8 lb 7.8 oz) ([DATE]) 4 %ile (Z= -1.70) based on WHO (Boys, 0-2 years) weight-for-age data using

9. WHO Growth Chart) Birth Weight: 2.48 kg (5 lb 7.5 oz) 2nd percentile wt-for-age Classification: SGA Change from Birthweight: 46% Current Weight: 3.63 kg (8 lb) ([DATE]) <1 %ile (Z= -2.86) based on WHO (Boys, 0-2 years) weight-for-age data using vitals from [DATE]. Current Length: 45 cm (17.72") ([DATE]) <1 %ile (Z= -4.02) based on WHO (Boys, 0-2 years) Length-for-age data based on Length recorded on [DATE]. Current HC: 35 cm (13.78") 7 %ile (Z= -1.49) based on WHO (Boys, 0-2 years) head circumference-for-age based on Head Circumference recorded on [DATE]. Wt/length: 92 %ile (Z= 1.38) based on

10. Change from Birthweight: 4% Birth Weight: 3.32 kg (7 lb 5.1 oz) ([DATE]) 3 %ile (Z= -1.90) based on WHO (Boys, 0-2 years) weight-for-age data using vitals from [DATE]. Classification: AGA Birth Length: 54 cm (21.26") ([DATE]) 98 %ile (Z= 2.10) based on WHO (Boys, 0-2 years) length-for-age data using vitals from [DATE]. Birth HC: 36 cm (14.17") 87 %ile (Z= 1.15) based on WHO (Boys, 0-2 years) head circumference-for-age data using vitals from [DATE]. Current Weight: 3.56 kg (7 lb 13.6 oz) ([DATE]) %ile (Z= -1.90) based on WHO (Boys, 0-2 years) weight-for-age data using

Query 1:

1. Dental varnish complete

2. Dental varnish complete.
3. Dental varnish complete.
4. Dental varnish complete.
5. Dental varnish complete.
6. Dental varnish completed
7. Dental varnish completed
8. Dental varnish applied
9. Dental varnish applied
10. Dental varnish applied

Query 2:

1. skill; Observed: a skill the examiner has observed directly; Reported: a skill that the caregiver reports, but which the examiner has not directly observed 18 month milestones Gross Motor Walks backward: negative Walks up stairs without help: negative Fine Motor Tower of 2 cubes: positive reported Language 6-20 words: positive reported Points to multiple named body parts: negative Attempts to combine words: negative Psychosocial Feeds self with cup and spoon (still messy): negative Tries to remove clothing: positive reported ,15 month Milestones Gross Motor Walks well: positive observed Walks up stairs with help: positive reported Stoops & recovers: positive reported

2. 15 Month Developmental Milestones for Parental Education: NOTE TO PARENTS: Most, but not all, children will begin to engage in some of the activities outlined below. What is most important is the progress your child makes in his or her ability to achieve these milestones over time. Gross Motor: Walks well; Walks up stairs with help; Stoops & recovers Fine Motor: Neat pincer grasp; Scribbles with crayon; Tower of 2 cubes Language: Mama and Dada (specific); 4 to 6 words; Points to 1 named body part Psychosocial:; Drinks from cup; Attempts use of spoon; Imitates housework; Tries to remove clothing

3. 15 Month Developmental Milestones for Parental Education: NOTE TO PARENTS: Most, but not all, children will begin to engage in some of the activities outlined below. What is most important is the progress your child makes in his or her ability to achieve these milestones over time. Gross Motor: Walks well; Walks up stairs with help; Stoops & recovers Fine Motor: Neat pincer grasp; Scribbles with crayon; Tower of 2 cubes Language: Mama and Dada (specific); 4 to 6 words; Points to 1 named body part Psychosocial:; Drinks from cup; Attempts use of spoon; Imitates housework; Tries to remove clothing

4. 15 Month Developmental Milestones for Parental Education: NOTE TO PARENTS: Most, but not all, children will begin to engage in some of the activities outlined below. What is most important is the progress your child makes in his or her ability to achieve these milestones over time. Gross Motor: Walks well; Walks up stairs with help; Stoops & recovers Fine Motor: Neat pincer grasp; Scribbles with crayon; Tower of 2 cubes Language: Mama and Dada (specific); 4 to 6 words; Points to 1 named body part Psychosocial:; Drinks from cup; Attempts use of spoon; Imitates housework; Tries to remove clothing

5. 15 Month Developmental Milestones for Parental Education: NOTE TO PARENTS: Most, but not all, children will begin to engage in some of the activities outlined below. What is most important is the progress your child makes in his or her ability to achieve these milestones over time. Gross Motor: Walks well; Walks up stairs with help; Stoops & recovers Fine Motor: Neat pincer grasp; Scribbles with crayon; Tower of 2 cubes Language: Mama and Dada (specific); 4 to 6 words; Points to 1 named body part Psychosocial:; Drinks from cup; Attempts use of spoon; Imitates housework; Tries to remove clothing

6. 15 Month Developmental Milestones for Parental Education: NOTE TO PARENTS: Most, but not all, children will begin to engage in some of the activities outlined below. What is most important is the progress your child makes in his or her ability to achieve these milestones over time. Gross Motor: Walks well; Walks up stairs with help; Stoops & recovers Fine Motor: Neat pincer grasp; Scribbles with crayon; Tower of 2 cubes Language: Mama and Dada (specific); 4 to 6 words; Points to 1 named body part Psychosocial:;

Drinks from cup; Attempts use of spoon; Imitates housework; Tries to remove clothing

7. 15 Month Developmental Milestones for Parental Education: NOTE TO PARENTS: Most, but not all, children will begin to engage in some of the activities outlined below. What is most important is the progress your child makes in his or her ability to achieve these milestones over time. Gross ,Motor: Walks well; Walks up stairs with help; Stoops & recovers Fine Motor: Neat pincer grasp; Scribbles with crayon; Tower of 2 cubes Language: Mama and Dada (specific); 4 to 6 words; Points to 1 named body part Psychosocial:; Drinks from cup; Attempts use of spoon; Imitates housework; Tries to remove clothing

8. 13 15 Month Developmental Milestones for Parental Education: NOTE TO PARENTS: Most, but not all, children will begin to engage in some of the activities outlined below. What is most important is the progress your child makes in his or her ability to achieve these milestones over time. Gross Motor: Walks well; Walks up stairs with help; Stoops & recovers Fine Motor: Neat pincer grasp; Scribbles with crayon; Tower of 2 cubes Language: Mama and Dada (specific); 4 to 6 words; Points to 1 named body part Psychosocial:; Drinks from cup; Attempts use of spoon; Imitates housework; Tries to remove

9. sub-score >=3 15 Month Developmental Milestones for Parental Education: NOTE TO PARENTS: Most, but not all, children will begin to engage in some of the activities outlined below. What is most important is the progress your child makes in his or her ability to achieve these milestones over time. Gross Motor: Walks well; Walks up stairs with help; Stoops & recovers Fine Motor: Neat pincer grasp; Scribbles with crayon; Tower of 2 cubes Language: Mama and Dada (specific); 4 to 6 words; Points to 1 named body part Psychosocial; Drinks from cup; Attempts use of spoon; Imitates housework; Tries to

10. has NOT acquired this skill; Observed: a skill the examiner has observed directly; Reported: a skill that the caregiver reports, but which the examiner has not directly observed) Gross Motor Walks backward: positive reported Walks up stairs without help: positive reported Kicks ball forward: positive reported Throws ball overhand: positive reported Fine Motor Tower of 2 cubes: positive reported Tower of 4 cubes: positive reported Dumps raisin/items: positive reported Language 6-20 words: negative Points to multiple named body parts: positive reported Attempts to combine words: negative Psychosocial Feeds self with cup and spoon (still messy): positive reported Tries to remove

Query 3: The following notes mean that immunization data are merged into patients' electronic health records.

1. Merged [State] Immunization Registry
2. Merged [State] Immunization Registry
3. Merged [State] Immunization Registry
4. Merged [State] Immunization Registry
5. Merged [State] Immunization Registry
6. Merged [State] Immunization Registry
7. Merged [State] Immunization Registry
8. Merged [State] Immunization Registry
9. Merged [State] Immunization Registry
10. Merged [State] Immunization Registry

Query 4:

1. Pediatrician
2. `[NAME]` last saw Pediatric Ophthalmology in [DATE]. They wanted to follow up in 6 months. No appointment scheduled. Referral placed for [NAME] to follow up with Pediatric Ophthalmology.
3. `[NAME]` missed his appointment with Pediatric Ophthalmology on [DATE]. Referral placed for `[NAME]` to follow up with Pediatric Ophthalmology.
4. Pediatric Hospital Medicine
5. `[NAME]` was last seen by Pediatric Ophthalmology in [DATE]. They wanted to follow up in 1 year. No appointment scheduled. Referral placed for [NAME] to follow up with Pediatric ophthalmology.

6. [NAME] last saw Pediatric Neurology in [DATE]. They wanted to follow up in 3 months. No appointment scheduled. Referral entered for [NAME] to follow up with Pediatric Neurology.

7. Pediatric Hospitalist

8. Pediatric Neurology

9. Pediatrics

10. Pediatrics

Query 5:

1. Esotropia, likely with some accommodative component - had improvement with glasses in past, however per mom, very poor glasses wear today Fixation preference OD, previous equal vision by PL OU, and last visit had OD two line better than OS but with poor attention - amblyopia suspect OS Anisometropia Stop atropine for now - will hopefully keep wearing glasses without it Use if needed, but no atropine within 2 weeks of the next visit Once nonaccommodative component of ET is clear, will likely need BMR rec

2. Esotropia, likely with some accommodative component - had improvement with glasses in past, however per mom, very poor glasses wear today Fixation preference OD, previous equal vision by PL OU, today OD two line better than OS but with poor attention Anisometropia Try atropine every day with glasses Looking nonaccommodative, so will schedule BMR rec re-measure at pre-op

3. 1. Intermittent alternating exotropia 2. Myopia of both eyes with regular astigmatism Visual Acuity (Toys) Right Left Dist sc CSM CSM Intermittent alternating exotropia - small angle exotropia seen today, intermittent and variable. Ortho at near. Mom thinks that both eyes drift. He has myopia with regular astigmatism. It is not significant enough to correct at this time. When he is older, if he remains myopic, glasses may help with the intermittent exotropia. I would monitor without glasses for now. No evidence of amblyopia at this time. Follow up 6 months with orthoptist.

4. Esotropia, likely accommodative component given improvement in glasses Keep wearing glasses full-time if possible (about 50% now) Anisometropia Possible refractive amblyopia - seeing well OU today, perhaps slightly better OD, monitor for now F/u 4 months with glasses, recheck alignment and vision with PL

5. Esotropia, likely with some accommodative component - had improvement with glasses in past Fixation preference OD, but equal VA by Teller cards New glasses equal to today's cycloplegic refraction Atropine OU once weekly to encourage glasses wear Anisometropia F/u 2 months Once we establish how much ET is accommodative, might need EOM surgery if residual ET

6. prescribe glasses at this time as patient is so young. Will probably need glasses once gets a little older for high myopia. Will reexamine in 3 months. If nystagmus is worse, will prescribe glasses at that time.

7. 1. Intermittent monocular esotropia of left eye 2. Suspected amblyopia of left eye Visual Acuity (Toys) Right Left Dist sc CSM CSUM Near sc CSM CSUM Small angle intermittent LET, prefers OD. Also with epicanthal folds which makes crossing appear worse. Discussed starting patching of OD, 1 hour a day, 5-6 days a week. No spectacles needed, minimal refractive error. Follow-up in 2 months with orthoptist, check VA and alignment. Amblyopia What is Amblyopia? When a young child uses one eye predominately and does not alternate between the two eyes, the prolonged suppression of the non-dominant eye by the brain

8. Esotropia, possible accommodative component Reviewed photographs brought by parents Give glasses prescription today with full CRx Anisometropia Possible refractive amblyopia - may need patching in the future F/u 6 weeks with glasses, recheck vision with PL

9. Esotropia Amblyopia suspect OD Hyperopia Glasses equal to today's cycloplegic refraction Follow-up in 3 months Might need patching

10. Goes by [NAME] Partially accommodative esotropia - angle a little less today No significant hyperopia Stop alternate patching Follow-up in 2 months If stable, will need surgery

Query 6:

1. teaching, evidenced no barriers to ,understanding, and showed good comprehension.
2. teaching, evidenced no barriers to ,understanding, and showed good comprehension.
3. receptive to teaching, evidenced no barriers to ,understanding, and showed good comprehension.
4. receptive to teaching, evidenced no barriers to ,understanding, and showed good comprehension.
5. receptive to teaching, evidenced no barriers to ,understanding, and showed good comprehension.
6. was receptive to teaching, evidenced no barriers to ,understanding, and showed good comprehension.
7. receptive to teaching, evidenced no barriers to understanding, and showed good comprehension.
8. receptive to teaching, evidenced no barriers to understanding and showed good comprehension.
9. receptive to teaching, evidenced no barriers to understanding and showed good comprehension.
10. receptive to teaching, evidenced no barriers to understanding and showed good comprehension.

Query 7:

1. started walking late, but [NAME] has made major improvement and mom no longer has physical development concerns. Per mom she very rarely hears him say words, but family members have told her that he does say words, one being "ma ma". Mom says he has a cousin his age that lives with him, and that has helped in his physical development but not with speech. Mom also states [NAME] acts like he doesn't hear sometimes because he completely ignores what you get him to do at times. During the assessment, [NAME] babbled here and there but did not say any
2. [NAME]'s parents first became concerned about his development at age 12-13 months when he did not respond to his name being called (although this has improved in the past 4-6 weeks). Communication: [NAME] began speaking in single words at age 10-11 months ("dada"). He currently speaks primarily by babbling (e.g., "mamama," "bababa"). He says "dada" but it is not directed towards his father. To indicate what he wants, he usually reaches and grunts. [NAME] is able to follow a few one-step directions. He understands "stop," "no," "come here" and sometimes "go to Dada." When his name is
3. URI sx. 3. Speech concern - has about 3-4 words: mama, dada, dog, quack, moo. Seems to understand basic commands but doesn't always respond. Responds to his name "when he wants to". 4. Bump below his lower lip for several days Older sister has autism Parents are seeing some similar signs He just recently started pointing, does have eye contact, has imaginative play Concerned about his poor response when talking to him And speech delay Per PCP note in [DATE]: · "Rapid weight gain - would cut back on 1 formula bottle · More protein and veggies for solids
4. Abnormal development Current concerns: Patient is demonstrating abnormal behaviors and regressions. Mother and father state that patient had initially started to babble around age 10 months and has since completely stopped speaking. Initially he would state mama and dada with intent but now does not indicate his needs. Father states that patient does not gesture and does not identify to family when he is hungry; they base his feeding schedule on when they typically eat. He frequently throws temper tantrums and tends to have sticky attention to objects or activities. When further asked, father states that he frequently plays repetitive
5. one day but not the next day. Developmental/Behavioral History: First Concerns: [NAME]'s parents first became concerned about his development at age 12-13 months when he did not respond to his name being called (although this has improved in the past 4-6 weeks). Communication: [NAME] be-

gan speaking in single words at age 10-11 months ("dada"). He currently speaks primarily by babbling (e.g., "mamama," "bababa"). He says "dada" but this is not directed towards his father. To indicate what he wants, he usually reaches and grunts. [NAME] is able to follow a few one-step directions. He understands "stop," "no," "come here" (in

6. Chief Complaint: Developmental concerns History of Present illness: [NAME] is a 12 m.o. male here for assessment of developmental delay. [NAME] was here recently for WCC and maternal concern for delay was also brought up at that time. Mother has noticed some repetitive behaviors where he seems to be fixated on whatever he is doing – specific examples include playing with spinning objects, playing with his cars - per mother he can have repetative motions where he will do the same behavior for long periods of time uninterrupted. He often times will not respond to his name and mother

7. for micrococcus species. Mom reports about a week prior to the increased clustering, she noticed him have increased falling and clumsiness. She believes his left side - both arm and leg - has been weaker as he will fall to that side. Mom is worried that is he starting to regress. She reports he can walk on his own and is working on leaning to run. He can climb stairs and will grab for his bottle with both hands. He can say "mama" and "dada". [NAME] has a history of infantile spasms and has been treated with Sabril, ACTH, high

8. that is he starting to regress. She reports he can walk on his own and is working on leaning to run. He can climb stairs and will grab for his bottle with both hands. He can say "mama" and "dada". ,[NAME] has a history of infantile spasms and has been treated with Sabril, ACTH, high dose steroids, and phenobarb. His neurologist had planned to start a ketogenic diet, however family was unable to because they lost medicaid coverage. Mom reports no seizure activity off of all medications from [DATE] to the end of [DATE] this year, when the current spells

9. 5 individual words, but now she only moans and cries. She can still walk fine, play fine, interact with others and make eye contact fine, but he vocabulary is now nonexistent. Mother states the child will still follow multiple-step commands, so there is no concerns about her hearing. The child was born full term vaginally with no complications, passed her newborn hearing test, did not have a lot of ear infections, met all her motor milestones on time, did not incur any traumatic injuries, and was speaking fine at her 15-month checkup. Growth charts have always been normal, including the

10. if this may have added to speech delay...had surgery in [DATE]. Gross Motor: [NAME] rolled over at 6 mo, sat unsupported at 14 mo (corrected age), not crawling yet and has not walked yet. Fine: [NAME] puts food in his mouth, transfers toys from hand to hand, no pincer grasp. Language: [NAME] was babbling by 14 mo (corrected), starting using words at 1 yr, put two words once "mama up" (14 mo corrected), and no sentences. Social: Doesn't always make eye contact but then sometimes makes really good eye contact; responds to music; calms when you sing to him;

B.2 Hyperpartisan Dataset

Query 0:

1. Hillary Clinton called Monica Lewinsky a “narcissistic loony tune.” She called Gennifer Flowers “some failed cabaret singer.” She said Republicans in Congress had organized “a vast right-wing conspiracy” against her husband, President Bill Clinton. And now the former secretary of state and defeated Democratic presidential candidate says those of us still interested in the investigating of her conduct are engaging in an “abuse of power.” Hillary Clinton expects special treatment. She always has. The real question is: why did the

2. entire conservative movement and much of the media. If the Clinton Foundation suffered a fraction of the problems of Trump’s, the media outrage would be deafening, yet not a single instance of quid pro quo has been unearthed. Seven Republican investigations into the Benghazi, Libya, terror attacks found nothing. So, they became probes of her emails, hoping to find something, anything, to pin

on her. Unfortunately for them, there was nothing incriminating, illegal, or untoward. Nothing. So now, during an

3. impulses and get away with it? Since the 1990s, how many men in powerful positions have seen Bill Clinton in that light? After all, all sorts of powerful people — from prominent feminists to powerful lawyers to the leaders of Clinton's party — came to the consensus that the whole Lewinsky mess was a "private matter." Perhaps the affair with her was — although Americans are right to expect better from a president — but the claims of Jones, Willey,

4. Then there's that pesky Uranium One scandal. Despite media efforts to portray the whole thing as conspiracy theory, there are still plenty of questions regarding the Russian money that flowed to the Clinton Foundation while Hillary was Secretary of State. In fact, there are enough questions that rumors of a special council persist - this despite Trey Gowdy's comments yesterday. So, the fine, upstanding, leftists over at Mother Jones decided to ask Mrs. Clinton about the possibility that she might

5. Hillary Clinton in last year's campaign. Liberals believe Comey gave the election to President Trump because her name was tarnished from Comey's actions but they totally disregard her criminal actions that led her to be in that place. Conservatives felt like Comey gave Hillary a pass for her obvious criminal actions. On June 27th Bill Clinton met with US Attorney General Loretta Lynch in her plane in Arizona for a half an hour. Within a week Lynch's Department of Justice

6. prosecutor. Don't let Washington ruin you, too. You need to send these Clintons back to where they came from." Read the full transcript below, and let us know what you think in the comments. Liberals Bash FBI Director Comey Over Clinton Probe After Praising Him in July Mother of Jailed Sailor: 'Hold Hillary to Same Standards' as My Son on Classified Info The FBI's ability to get Huma Abedin- Hillary Clinton's closest advisor, confidante and State Department deputy chief of

7. step into politicizing the Justice Department" and "such an abuse of power." In an exclusive interview with Mother Jones, Clinton said such an investigation would have devastating consequences for the justice system in America. "If they send a signal that we're going to be like some dictatorship, like some authoritarian regime, where political opponents are going to be unfairly, fraudulently investigated, that rips at the fabric of the contract we have, that we can trust our justice system," Clinton said.

8. Only hours after President Donald Trump's aide, Kellyanne Conway, had a fiery exchange with CNN's Chris Cuomo over whether or not the White House was obsessed about Hillary Clinton, the president completely undercut her argument. On Thursday morning, the president, spending his " Executive Time" watching "Fox & Friends," angrily tweeted about his Democratic opponent in 2016, while slamming the infamous dossier in an attempt to distract from the ongoing probe by special counsel Robert Mueller into the Trump campaign's

9. leaked and totally protected Hillary Clinton. He was the best thing that ever happened to her!" In interview excerpts released in August by the Senate Judiciary Committee, FBI officials said Comey and investigators had determined by the spring of 2016 that charges weren't warranted, and had begun thinking of how the public should be informed of that decision. Clinton was interviewed by the FBI in early July, just days before Comey announced the investigation's conclusion. That timing has prompted criticism

10. was a lot of work going on behind the scenes. Given the security level, it's likely much of the information available at the time could not be made public. She noted she wished Obama had acted prior to the election but asked what Trump is doing today to stop Russia from acting in the 2018 election. The Fox host along with Pavlich went on to claim that Clinton was colluding with Russia because her campaign partially paid the law firm

Query 1:

1. have yet acknowledged their bygone failures of imagination, or granted that civil libertarians were right: The establishment has permitted the American presidency to get dangerously powerful. While writing or sharing articles that compare Trump to Hitler, Mussolini, and Franco, few if any have called on Obama or Congress to act now "to tyrant-proof the White House." However much they fear Trump, however rhetorically maximalist their warnings, even the prospect of him controlling the national security state does not cause

2. is so destructive because his enemies help him. He ramps up the aggression. His enemies ramp it up more, to preserve their own dignity. But the ensuing cultural violence only serves Trump's long-term destructive purpose. America is seeing nearly as much cultural conflict as it did in the late 1960s. It's quite possible that after four years of this Trump will have effectively destroyed the prevailing culture. The reign of the meritocratic establishment will be just as over as the

3. It has to be admitted that Donald Trump is doing exactly what he was elected to do. He was not elected to be a legislative president. He never showed any real interest in policy during the campaign. He was elected to be a cultural president. He was elected to shred the dominant American culture and to give voice to those who felt voiceless in that culture. He's doing that every day. What's troubling to me is that those who are

4. Obama, Trump is a natural, predictable endpoint. Furthermore, Trump is what happens when you wear your Christian conservative values like a cardigan to conveniently slip off when the heat rises. Trump is fundamentally altering American politics — coarsening them, corrupting them, cratering them. And America, particularly conservative America, has only itself to blame. Republicans sowed intolerance and in its shadow, Trump sprang up like toxic fungi.

5. President Donald Trump speaks during a meeting with Governor Ricardo Rossello of Puerto Rico in the Oval Office of the White House, Thursday, Oct. 19, 2017, in Washington. (AP Photo/Evan Vucci) The lobotomization of the Republican Party appeared complete last year when the same GOP paladins who had denounced Donald Trump as a "lunatic trying to get ahold of nuclear weapons" (Marco Rubio), as a bigot who was guilty of "the textbook definition of a racist comment" (Paul Ryan), and

6. "It is painfully obvious that the communism lovers in our institutions of higher education have had the run of the asylum and wreaked havoc." Many liberal talking heads have made a big deal out of the supposed demographic wave facing Republicans and conservatives as the nation morphs into a more diverse population. Growth rates amongst minorities have begun to exceed those of whites. The narrative goes like this — as the white, European-based population diminishes, America will become less conservative

7. borders to prohibit illegal entry, especially entry by dreadful thugs and terrorists, and moving aggressively to rid the country of those thugs and illegals like MS-13 and the monster who killed Kate Steinle, whom Obama welcomed, either directly or through insolent benign neglect. And on and on. We are awestruck at the vision, purpose and energy Trump has brought to the leadership of his country. But he is yet hampered by a war at home. The Democrats, the Progressives (or

8. at an annual rate of 3 percent. All those shibboleths have either been blown up or may yet be blown up in 2018. Trump is no longer written off by the Left as a sleepy dud. Instead, he suddenly is being redefined by many of his progressive enemies as a dangerous workaholic and right-wing revolutionary. Never-Trump Republicans no longer insist that Trump is a liberal Manhattan wolf in conservative sheep's clothing. Grudgingly, they now confess that he is ramming through

9. his administration. No need to replay senior members of his own party denouncing his behavior. It's become glaringly obvious: this guy is about one thing, and that is being the stand-in for a basket of Americans once labeled "deplorables". While this sounds like a petty insult, the polls have been consistent:

approximately 1/3rd of Americans have deplorable views about black people, people of color, LGBT, and immigrants. They are overwhelmingly white, and while they are not all Trump supporters, ALL