

1A. AI use case name	1C. Office with AI use case	Provide a short summary (200 words max) of what the AI does.
Health Resources and Services Administration (HRSA) Electronic Handbooks (EHBs) AI Chatbot	Health Resources and Services Administration (HRSA)	<p>AI Chatbot</p> <ul style="list-style-type: none"> • Successfully developed and deployed HRSA EHBs AI Chatbot using Artificial Solutions Teneo platform for external HRSA EHBs grantees • Built to allow grantees to communicate with the EHBs Chatbot using regular natural conversational expressions • Provides knowledge- and action-based responses through a self-service platform with 24/7 availability • Integrated with existing EHBs application UI and Salesforce for automated ticket creation • Chatbot has the ability to refine and increase the accuracy of its responses as more and more users invoke/use the Chatbot
Health Resources and Services Administration (HRSA) BHW Community Need Analysis Platform	Health Resources and Services Administration (HRSA)	The first use case being developed is for primary care with behavioral health integration which uses a machine learning based automated clustering engine. The development of this tool allows for BHW to dynamically assess the healthcare need of a population given a specific use case and relevant datasets. The output of the model will be used as part of the Notice of Funding Opportunity (NOFO) grant proposal evaluation process.
ICD-10 Coding of Cause of Death reported on Death Certificates (MedCoder)	Centers for Disease Control and Prevention (CDC) NCHS	MedCoder ICD-10 cause of death codes to the literal text cause of death description provided by the cause of death certifier on the death certificate. This includes codes for the underlying and contributing causes of death.
Item Nonresponse Detection in Open-text Response Data	Centers for Disease Control and Prevention (CDC) NCHS	NCHS is developing an item nonresponse detection model, to identify cases of item nonresponse (e.g., gibberish, uncertain/don't know, refusals, or high-risk) among open-text responses to help improve survey data and question and questionnaire design. The system is a Natural Language Processing (NLP) model pre-trained using Contrastive Learning and fine-tuned on a custom dataset from survey responses.
Sequential Coverage Algorithm (SCA) in Record Linkage	Centers for Disease Control and Prevention (CDC) NCHS	CDC's National Center for Health Statistics (NCHS) Data Linkage Program has implemented a supervised machine learning algorithm, known as the Sequential Coverage Algorithm (SCA) in their linkage programs. The SCA was used to develop joining methods (or blocking groups) when working with very large datasets. The SCA method improved the efficiency of blocking.
Chatbot - Voice	Centers for Medicare & Medicaid Services (CMS)	CMS/OSFLO: To assist the CMS Badging Help Desk, this Chatbot (voice) is an automated phone response for general badging questions allowing help desk personnel to assist employees and contractors with more detailed/larger issues.
Chatbot - Text	Centers for Medicare & Medicaid Services (CMS)	CMS/OSFLO: To assist the Security team, this Chatbot (text) provides an automated email response for general physical security questions, allowing the help desk team to assist employees and contractors with more in depth issues.
Feedback Analysis Solution (FAS)	Centers for Medicare & Medicaid Services (CMS)	The Feedback Analysis Solution is a system that uses CMS or other publicly available data (such as Regulations.Gov) to review public comments and/or analyze other information from internal and external stakeholders. The FAS uses Natural Language Processing (NLP) tools to aggregate, sort and identify duplicates to create efficiencies in the comment review process. FAS also uses machine learning (ML) tools to identify topics, themes and sentiment outputs for the targeted dataset.
Predictive Intelligence - Incident Assignment for Quality Service Center (QSC).	Centers for Medicare & Medicaid Services (CMS)	Predictive Intelligence (PI) is used for incident assignment within the Quality Service Center (QSC). The solution runs on incidents created from the ServiceNow Service Portal (https://cmsqualitysupport.servicenowservices.com/sp_ess). The solution analyzes the short description provided by the end user in order to find key words with previously submitted incidents and assigns the ticket to the appropriate assignment group. This solution is re-trained with the incident data in our production instance every 3-6 months based on need.
Reasonable Accommodation RPA Bot	Centers for Medicare & Medicaid Services (CMS)	The Bot pulls HR data related to staffing changes, e.g. promotions, reassignments, change in supervisor, and generates information for action by Reasonable Accommodation staff to ensure disability reasonable accommodations follow the employee.
Rapid Authority to Operate (ATO)	Centers for Medicare & Medicaid Services (CMS)	The Rapid ATO System was built using a natural language processing model and pipeline to process system security plans, to identify unique and commonly used technology components used across Federal Information Security Management Act (FISMA) systems. Natural language processing (NLP) is a form of machine learning that derives intent or subject out of blocks of text. In this particular case it was used to identify common blocks of language used in similar ways across system security plan (SSP) documents. In this way, CMS could identify similar approaches to solving certain technology or process-related control areas within the Acceptable Risk Safeguards (ARS). The output was used to create a list of components to develop control description language in a re-usable way, as part of the Blueprint/Rapid ATO effort to streamline SSP generation for new systems.
Data Lake/Load-Extract-Load-Transform (L-ETL)	Centers for Medicare & Medicaid Services (CMS)	CMS is using Security Data Lake to modernize the load-extract-load-transform (L-ETL) pipelines and data tooling. CMS will be enhancing Agency security to bring together more system, telemetry and program data in one place with a unifying governance model. Building on top of a modern data platform will provide opportunities to experiment with machine learning model development against this data--solving any number of problems that require decisions to be made about inferences over time series data. There is no actual ML/AI work being done here today, rather, we are beginning work on the scaffolding that will open up these opportunities in 1-2 years time.
Priority Score Model - ranks providers within the Fraud Prevention System using logistic regression based on program integrity guidelines.	Centers for Medicare & Medicaid Services (CMS)	Inputs - Medicare Claims data, Targeted Probe and Educate (TPE) Data, Jurisdiction information Output - ranks providers within the FPS system using logistic regression based on program integrity guidelines.
Priority Score Timeliness - forecast the time needed to work on an alert produced by Fraud Prevention System (Random Forest, Decision Tree, Gradient Boost, Generalized Linear Regression)	Centers for Medicare & Medicaid Services (CMS)	Inputs - Medicare Claims data, TPE Data, Jurisdiction information Output - forecast the time needed to work on an alert produced by FPS (Random Forest, Decision Tree, Gradient Boost, Generalized Linear Regression)
Provider Education 90 Day - reviews claims for provider before and after education for statistical change in their claim submission patterns	Centers for Medicare & Medicaid Services (CMS)	Inputs - Medicare Claims data, TPE Data, Jurisdiction information Output - reviews claims for provider before and after education for statistical change in their claim submission patterns
Advanced Semantic Search and Indexing of Text for Tobacco Applications (ASSIST4Tobacco)	Food and Drug Administration (FDA)	ASSIST4Tobacco is a novel tool that will use semantic indexing to search tobacco authorization applications. The system will be based on an AI (artificial intelligence)-based NLP (Natural Language Processing) model which provides deeper search capabilities using a language model developed to represent relationships between words and concepts within a body of text.
Artificial Intelligence-based Deduplication Algorithm for Classification of Duplicate Reports in the FDA Adverse Event Reports (FAERS)	Food and Drug Administration (FDA)	The deduplication algorithm is applied to nonpublic data in the FDA Adverse Event Reporting System (FAERS) to identify duplicate reports. Unstructured data in free text FAERS narratives is processed through a natural language processing system to extract relevant clinical features. Both structured and unstructured data are then used in a probabilistic record linkage approach to score pairs of reports by evaluating multiple data fields and applying relative weights per field. The output of potential duplicate reports is further placed in groups to facilitate identification of FAERS reports during case series evaluation for safety issues of concern.

1A. AI use case name	1C. Office with AI use case	Provide a short summary (200 words max) of what the AI does.
Opioid Data Warehouse Term Identification and Novel Synthetic Opioid Detection and Evaluation Analytics	Food and Drug Administration (FDA)	The Term Identification and Novel Synthetic Opioid Detection and Evaluation Analytics use publicly available social media and forensic chemistry data to identify novel referents to drug products in social media text. It uses the FastText library to create vector models of each known NSO-related term in a large social media corpus, and provides users with similarity scores and expected prevalence estimates for lists of terms that could be used to enhance future data gathering efforts.
National Institute of General Medical Sciences (NIGMS) AI Supported Searches, Information Systems and Tools System Acronym: NIGMS ASSIST	National Institutes of Health (NIH) NIGMS	<p>NIGMS program staff often need information that is available through IMPAC or QVR to perform their daily tasks. In order to provide such information, DIMA and IRMB have collaborated to develop functions that utilize artificial intelligence and natural language processing methods to produce data relevant to the program staff's mission. These tools are collected into a single system to make them available to the NIGMS community for use on a day-to-day basis. ASSIST provides a secure interface supported by Oracle, SQL server and Python analytics. The individual components of ASSIST provide the following functions:</p> <ul style="list-style-type: none"> - FLIP module (Development), provides the ability to identify investigators by PPID from Federal RePORTER based on user input of investigator PPIDs. - TPAL module (Production), provides the ability to lookup potential matching program officers, including their corresponding predicted Program Area Codes, and ICs based on the input of unstructured scientific data.
Leveraging AI for Business Process Automation	National Institutes of Health (NIH) NIGMS	NIGMS has developed a method to automate the initial referral of grant applications to the proper scientific expertise within the Institute using Natural Language Processing and Machine Learning. NIGMS IRMB and DIMA are currently using this NLP/ML algorithm developed in R statistical software to parse grant applications and to determine Project Officer candidates for grant assignment. This process was previously fully manual and required a substantial person hour effort. NIGMS has collaborated with the Electronic Records Administration group to incorporate this technique into the Internal Referral Module, and the tool is now available to be adapted for broader use across the NIH.
Grant Application Subject-Matter Classification Tool	National Institutes of Health (NIH) NIEHS	Natural language processing of grant applications for the purpose of classification for review assignment
Splunk IT System Monitoring Software	National Institutes of Health (NIH) NIEHS	Splunk utilizes machine learning to aggregate system logs from IT infrastructure systems and endpoints for auditing and monitoring purposes
COVID-19 Pandemic Vulnerability Index Dashboard	National Institutes of Health (NIH) NIEHS	The dashboard creates risk profiles, called PVI scorecards, for every county in the United States, continuously updated with the latest data that summarize and visualize overall disease risk.
Leveraging AI/ML for classification and categorization of scientific concepts	National Institutes of Health (NIH) NHLBI	Topical characterization of the research portfolio. Inputs are publications and grants abstracts. These are fed into a text classification model and concept extraction. The outputs are category labels and list of concepts.
Machine learning system to predict translational progress in biomedical research	National Institutes of Health (NIH) OPA	<p>Fundamental scientific advances can take decades to translate into improvements in human health. Shortening this interval would increase the rate at which scientific discoveries lead to successful treatment of human disease. One way to accomplish this would be to identify which advances in knowledge are most likely to translate into clinical research. Toward that end, the NIH Office of Portfolio Analysis built a machine learning system that detects whether a paper is likely to be cited by a future clinical trial or guideline. Despite the noisiness of citation dynamics, as little as 2 years of postpublication data yield accurate predictions about a paper's eventual citation by a clinical article (accuracy = 84%, F1 score = 0.56; compared to 19% accuracy by chance). We found that distinct knowledge flow trajectories are linked to papers that either succeed or fail to influence clinical research. Translational progress in biomedicine can therefore be assessed and predicted in real time based on information conveyed by the scientific community's early reaction to a paper. For more information see the publication describing this system: Hutchins et al 2019 (https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3000416)</p>
Semantic analysis of scientific documents (word2vec _{OPA})	National Institutes of Health (NIH) OPA	The NIH Office of Portfolio Analysis has developed a neural network approach to analysis of scientific content using dimensionality reduction (word2vec _{OPA}). This method computationally converts words in scientific texts to numbers and summarizes documents by their semantic content by learning relationships between words from their context. This method is adaptable to specific corpora, including grants and scientific articles. For more information see the publication describing our word2vec approach: Hoppe et al 2019 (https://www.science.org/doi/10.1126/sciadv.aaw7238)
Person-level disambiguation for PubMed authors and NIH grant applicants	National Institutes of Health (NIH) OPA	<p>High-quality disambiguation is required to correctly link researchers to their grants and outputs including articles, patents, and clinical trials. The NIH Office of Portfolio Analysis developed a disambiguation solution that used article level metadata to assign 24.5M unique papers from the PubMed database to 16.0M unique author names, then used a novel neural network model trained on ORCID identifiers to determine whether author-publication pairs refer to variant representations of the same person. For example, our model can determine whether hypothetical records listing Jane Smith and Jane M. Smith were the same person, or two different people, based on variables that include institutional affiliation, co-authorship, and article-affiliated Medical Subject Heading (MeSH) terms. For more information see the publication describing this method: Yu et al 2021 (https://www.biorxiv.org/content/10.1101/2021.02.02.429450v1.full.pdf)</p>
Program Class Code (Area of Science) Referral for NIAID	National Institutes of Health (NIH) OER	The REFERRAL GROUP of Referral, Program Analysis Branch (RPAB) is responsible for program assignments for all research, training, career, and fellowship grant applications submitted to NIAID, from CSR. The NIAID Program Class Code classification AI project evaluates the projects that are in RAPB and auto assigns these grant applications to the Program Class Codes. The inputs are comprised of approximately 6,000+ grant applications that are currently manually assigned by RPAB Staff. The output would be grant applications that are categorized into their respective PCC's.
Research, Condition, and Disease Categorization (RCDC)	National Institutes of Health (NIH) OER	RCDC is an electronic budget reporting tool that categorizes projects using AI/NLP. The inputs are grant applications, R&D contracts, intramural projects, inter agency agreements. The RCDC Fingerprinting process identifies concepts in the extracted text from the source project, person or publication. The text is normalized, concepts are extracted, concepts and synonyms are matched to the RCDC thesaurus. A rank is applied based on the frequency of occurrence of the concepts within the text. Project fingerprints are sourced from the application description text (Title, Abstract and Specific Aims). Titles and abstracts provide the source of scientific concepts for publications. The system then outputs the projects into their respective areas of science.
Query View Report (QVR) LIKE	National Institutes of Health (NIH) OER	The LIKE feature in QVR makes use of the NIH Research, Condition and Disease Categorization (RCDC) indexing results to compare scientific terms associated with a project, person or publication and find scientifically similar projects, persons or publications.
Internal Referral Module (IRM) NLP	National Institutes of Health (NIH) OER	The IRM NLP module automatically refers projects to Program Officers once the grant application is received. The system inputs are grant applications - the title, abstract, specific aims and Public Health Relevance is analyzed to automatically refer the grant application to the Program Officer who matches a similar background with the science contained in the applications. This process, is operating at a high accuracy rate and has effectively eliminated the referral bottleneck.
NIH Grants Virtual Assistant	National Institutes of Health (NIH) OER	Chat Bot to assist users in finding grant related information via OER resources

1A. AI use case name	1C. Office with AI use case	Provide a short summary (200 words max) of what the AI does.
Pangolin lineage classifications to support accessing and analysis of SARS-CoV-2 sequence data.	National Institutes of Health (NIH) NLM	The Pango nomenclature, called Pango lineages, is being used by researchers and public health agencies worldwide to track the transmission and spread of SARS-CoV-2, including variants of concern. The requirements for running the tool include having conda on a MacOS or Linux system, and the FASTA-formatted sequence data. There are 2 methods for lineage assignment with Pango; within NCBI Virus we use the process which includes PangoLEARN, where a classification tree is used to group similar sequences.
Providing MeSH Check Tag of NLM's Medical Text Indexer (MTI) ons using Support Vector Machines (SVM)	National Institutes of Health (NIH) NLM	Titles and abstracts from MEDLINE Citations are provided through SVM machine learning algorithm provides confidence scores for a set of MeSH CheckTags to the NLM Medical Text Indexer (MTI) program. These CheckTags are small set of MeSH Descriptors designed to indicate Species, Sex, and Age in MEDLINE articles.
Determining selection for indexing MEDLINE articles using Neural Network Architecture with a Convolutional Neural Network (CNN)	National Institutes of Health (NIH) NLM	Using the article title, abstract, journal, publication year, and indexing year of indexed and non-indexed articles that were submitted to MEDLINE in 2018, methods to automate the selection of indexed articles was researched. A classifier was developed that combines the predictions of many traditional machine learning algorithms and a Convolutional Neural Network (CNN). The final classification layer uses a sigmoid activation function to generate a single output value between zero and one, which can be interpreted as the probability of an article being in-scope for MEDLINE.
MetaMap to identify potential terms for indexing MEDLINE articles	National Institutes of Health (NIH) NLM	MetaMap is a widely available program providing access from biomedical text to the concepts in the unified medical language system (UMLS) Metathesaurus. MetaMap uses NLP to provide a link between the text of biomedical literature and the knowledge, including synonymy relationships, embedded in the Metathesaurus. The flexible architecture in which to explore mapping strategies and their application are made available. MTI uses the MetaMap to generate potential indexing terms.
Best Match: New relevance search for PubMed	National Institutes of Health (NIH) NLM	PubMed is a free search engine for biomedical literature accessed by millions of users from around the world each day. With the rapid growth of biomedical literature, finding and retrieving the most relevant papers for a given query is increasingly challenging. We have developed Best Match, a new relevance search algorithm for PubMed that leverages the intelligence of our users and cutting-edge machine-learning technology as an alternative to the traditional date sort order. The Best Match algorithm is trained with past user searches with dozens of relevance-ranking signals (factors) and demonstrates state-of-the-art retrieval performance in benchmarking experiments as well as an improved user experience in real-world testing.
SingleCite: Improving single citation search in PubMed	National Institutes of Health (NIH) NLM	A search that is targeted at finding a specific document in databases is called a Single Citation search, which is particularly important for scholarly databases, such as PubMed, because it is a typical information need of the users. We have developed SingleCite, an automated algorithm that establishes a query-document mapping by building a regression function to predict the probability of a retrieved document being the target based on three variables: the score of the highest scoring retrieved document, the difference in score between the two top retrieved documents, and the fraction of a query matched by the candidate citation. SingleCite shows superior performance in benchmarking experiments and is applied to rescue queries that would fail otherwise.
Computed Author: author name disambiguation for PubMed	National Institutes of Health (NIH) NLM	PubMed users frequently use author names in queries for retrieving scientific literature. However, author name ambiguity (different authors share the same name) may lead to irrelevant retrieval results. Thus we have developed a machine-learning method to score the features for disambiguating a pair of papers with ambiguous names. Subsequently, agglomerative clustering is employed to collect all papers belong to the same authors from those classified pairs. Disambiguation performance is evaluated with manual verification of random samples of pairs from clustering results, with a higher accuracy than other state-of-the-art methods. It has been integrated into PubMed to facilitate author name searches.
National Library of Medicine NLM-Gene: towards automatic gene indexing in PubMed articles	National Institutes of Health (NIH) NLM	Gene indexing is part of the NLM's MEDLINE citation indexing efforts for improving literature retrieval and information access. Currently, gene indexing is performed manually by expert indexers. To assist this time-consuming and resource-intensive process, we have developed NLM-Gene, an automatic tool for finding gene names in the biomedical literature using advanced natural language processing and deep learning methods. Its performance has been assessed on gold-standard evaluation datasets and is to be integrated into the production MEDLINE indexing pipeline.
National Library of Medicine NLM-Chem: towards automatic chemical indexing in PubMed articles	National Institutes of Health (NIH) NLM	Chemical indexing is part of the NLM's MEDLINE citation indexing efforts for improving literature retrieval and information access. Currently, chemical indexing is performed manually by expert indexers. To assist this time-consuming and resource-intensive process, we have developed NLM-Chem, an automatic tool for finding chemical names in the biomedical literature using advanced natural language processing and deep learning methods. Its performance has been assessed on gold-standard evaluation datasets and is to be integrated into the production MEDLINE indexing pipeline.
Grants Analytics Portal	Office of Inspector General (OIG)	The Grants Analytics Portal uses AI to enhance HHS OIG staff's ability to access grants related data quickly and easily by: quickly navigating directly to the text of relevant findings across thousands of audits, the ability to discover similar findings, analyze trends, compare data between OPDIVs, and the means to see preliminary assessments of potential anomalies between grantees.
Text Analytics Portal	Office of Inspector General (OIG)	The text analytics portal allows personnel without an analytics background to quickly examine text documents through a related set of search, topic modeling and entity recognition technologies
Relevancy Tailoring	Agency for Healthcare Research and Quality (AHRQ)	Adjusting the ranking of search results so that most relevant results show up at the top of the list
Auto-generation Synonyms	Agency for Healthcare Research and Quality (AHRQ)	Behind the scenes, adding synonyms to search queries to improve search results
Automated Suggestions	Agency for Healthcare Research and Quality (AHRQ)	Auto-filling queries as they are typed
Suggested Related Content	Agency for Healthcare Research and Quality (AHRQ)	Show related searches that may provide the user with other related, valuable information
Auto Tagging	Agency for Healthcare Research and Quality (AHRQ)	Suggesting content tags automatically based on a machine-driven evaluation of how existing content is tagged
Did you mean	Agency for Healthcare Research and Quality (AHRQ)	Suggesting spelling corrections and reformatted search queries based on Google Analytics data
Chatbot	Agency for Healthcare Research and Quality (AHRQ)	An interactive interface that can respond to plain language queries in real time using natural language processing