**How to Fine-Tune GPT-3 Model for Named Entity Recognition**

**Introduction**

Fine-tuning is a critical process that allows you to customize a pre-trained model to better suit your specific application. With access to the vast amount of data that has been used to train models like GPT-3, fine-tuning empowers you to improve the results you get from these models for your unique use case.

One of the most significant advantages of fine-tuning is that it enables “few-shot learning,” whereby the model can intuit the task you are trying to perform based on just a few examples. This process goes even further by allowing you to train on a much larger set of examples than what typically fits in a prompt, resulting in even more exceptional performance.

The process of fine-tuning a model can be broken down into several fundamental steps. Firstly, you need to prepare and upload your training data, which should be specific to your application and representative of the task you want the model to perform. Once the data is uploaded, you can train a new fine-tuned model, a process that may take time, but is well worth the effort, as it results in a model that is much better suited to your specific needs.

Once your fine-tuned model is trained, you can start using it to generate superior results. Importantly, once a model has been fine-tuned, you no longer need to provide examples in the prompt, resulting in cost savings and enabling lower-latency requests.

Fine-tuning is a powerful technique that can help you achieve better results on a wide range of tasks, while also saving you costs in the long run. By capitalizing on the potential of pre-trained models like GPT-3, fine-tuning can play an instrumental role in elevating your organization’s performance.

In this article, we will show how to prepare the data to fine-tune a GPT-3 model on Named Entity Recognition (NER) and run inference.

**Preparing Training Data**

Preparing the training data is an essential step in fine-tuning GPT-3, as it teaches the model the desired output. The data must be in the form of a JSONL document, where each line is a prompt-completion pair corresponding to a training example. This format is easy to work with, and you can use the CLI data preparation tool to convert your data into this format quickly.

When designing your prompts and completions for fine-tuning, it is essential to remember that it is different from prompt engineering . For example, when using the base model, prompts often consist of multiple concatenated examples, but for fine-tuning, each training example generally consists of a single input example and its associated output. This eliminates the need to give detailed instructions or include multiple examples in the same prompt reducing the cost significantly.

To ensure the best results, it is recommended to have at least a 100 examples. The more training examples you have, the better the model will perform. In general, it is found that each doubling of the dataset size leads to a linear increase in model quality up to a certain point.

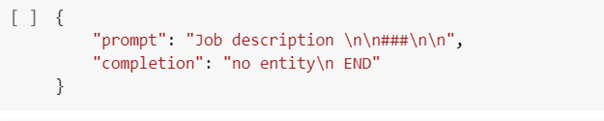
In summary, preparing training data is a crucial step in fine-tuning GPT-3, and it is essential to have a good understanding of how to design your prompts and completions. With the right training data, you can achieve excellent results and customize GPT-3 to suit your specific needs

To fine-tune a model, you’ll need a set of training examples that each consist of a single input (“prompt”) and its associated output (“completion”):

* The data formatting is a crucial step, and it’s important to make sure that your prompt and completion follows the correct format. The prompt should end with a **fixed separator** (e.g. \n\n###\n\n) to inform the model when the prompt ends and the completion begins. Similarly, the completion should start with a **whitespace** and end with a fixed stop sequence (e.g. \n or ###) to inform the model when the completion ends.

{"prompt":"This position requires that you: Have a BS in Computer Science, related degree, or equivalent experience Have 2+ years of experience in the technology industry Can work with C++ and\/or Objective-C code Are able to learn other frameworks, such as ReactNative\n\n###\n\n","completion":" diploma:['bs']\ndiploma\_major:['computer science']\nexperience:['2 + years']\nskills:['technology industry', 'c++', 'objective - c']\n END"}

* When preparing your dataset, it’s important to remember that the separator, stop sequence and whitespace should not appear elsewhere in any prompt or completion. This will help the model to understand the boundaries between the prompt and completion and produce accurate results.
* It is important to provide negative examples so the model knows where there are no entities to extract and doesn’t “hallucinate” or fabricate an answer.



Training example with no entities

* For inference, it’s important to format your prompts in the same way as you did when creating the training dataset, including the same separator and stop sequence. This will allow the model to properly truncate the completion.

**Problem setup**

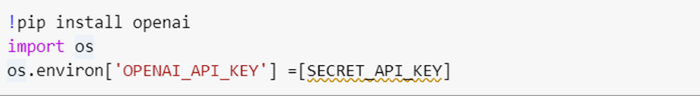
Our objective is to develop an NER system catered to job descriptions and resumes that can accurately identify and extract targeted entities from the text. The dataset we intend to employ consists of 75 descriptions that have been categorized into 4 distinct terms: DIPLOMA, DIPLOMA\_MAJOR, EXPERIENCE, and SKILLS.

**GPT-3 Fine-tuning**

To fine-tune our model, we must first build a training dataset that meets the specifications of the OpenAI API. This requires us to provide a JSONL file containing examples separated by line breaks, each formatted as follows:

{"prompt":"Required Skills and Experience: At least 5 years of relevant work experience in a Datacenter or other critical environment facility, working with the operation of building cooling, electrical, mechanical and life safety systems, Electrical Power Monitoring Systems (EPMS), Branch Circuit Monitoring (BCM), Building Automation Systems (BAS), and Battery Monitoring Systems (BMS) At least three years experience leading and motivating a diverse, technical workforce Preferred Skills and Experience : Bachelors Degree or Technical College certification in mechanical or electrical engineering and\/or services Experience working on large scale CE projects Experience with the operation of IT infrastructure (Servers, SANs, Networking, etc.) Background Check Requirements: Ability to meet Microsoft, customer and\/or government security screening requirements are required for this role. These requirements include, but are not limited to the following specialized security screenings: Citizenship Verification: This position requires verification of US Citizenship to meet federal government security requirements. Candidates must have an active TS and be willing to upgrade to TS\/SCI (with polygraph) or have an active TS\/SCI and be willing to upgrade to TS\/SCI (with polygraph). This role will require candidates to maintain the TS\/SCI (with polygraph) clearance. Microsoft Cloud Background Check: This position will be required to pass the Microsoft Cloud background check upon hire\/transfer and every two years thereafter.\n\n###\n\n","completion":" experience:['5 years', 'three years']\nskills:['datacenter', 'critical environment facility', 'operation of building cooling', 'electrical', 'mechanical', 'life safety systems', 'electrical power', 'monitoring systems ( epms )', 'branch circuit monitoring ( bcm )', 'building automation systems ( bas )', 'battery monitoring systems ( bms )', 'leading', 'motivating']\ndiploma:['bachelors']\ndiploma\_major:['mechanical', 'electrical engineering']\n END"}  
{"prompt":"This positionrequires that you: Have a BS in Computer Science, related degree, or equivalent experience Have 2+ years of experience in the technology industry Can work with C++ and\/or Objective-C code Are able to learn other frameworks, such as ReactNative\n\n###\n\n","completion":" diploma:['bs']\ndiploma\_major:['computer science']\nexperience:['2 + years']\nskills:['technology industry', 'c++', 'objective - c']\n END"}  
{"prompt":"Bachelors degree in Computer Science or related field, or equivalent industry experience 5+ years of experience delivering solutions and support to enterprise customers 2+ years of experience managing and leading highly technical teams in a fast-paced environment. Demonstrated hands-on experience on one or more of the Dynamics 365 products e.g. Dynamics Customer Engagement (CRM), Dynamics Finance & Operations (ERP) Preferred: MBA Understanding of cloud computing technologies is desired - Azure Core Platform; Data Platform: SQL, Azure DB ; Application development & debugging experience; Power BI, PowerApps Strong passion and focus on delivering the right customer experience Demonstrated ability to recruit and develop global teams Ability to innovate and drive change Ability to build a deep technical relationship with internal teams and customers Microsoft Cloud Background Check: This position will be required to pass the Microsoft Cloud background check upon hire\/transfer and every two years thereafter. Leave this section at the bottom:\n\n###\n\n","completion":" diploma:['bachelors']\ndiploma\_major:['computer science']\nexperience:['5 + years', '2 + years']\nskills:['delivering solutions', 'managing', 'leading highly technical teams']\n END"}  
{"prompt":"5yrs experience testing in complex systems (OS, Virtualization, Storage, or Networking products). This experience includes writing test plans, developing tests, test process Preferred Tech and Prof Experience While one of the areas of testing are , the more of (Operating System, Networking, storage, virtualization) they have the better. Familiarity and experience with scripting languages, Python, and tool development also add to strength of resume. Experience with agile development process is also preferred. EO Statement IBM is committed to creating a diverse environment and is proud to be an equal opportunity employer. All qualified applicants will receive consideration for employment without regard to race, color, religion, gender, gender identity or expression, sexual orientation, national origin, genetics, disability, age, or veteran status. IBM is also committed to compliance with all fair employment practices regarding citizenship and immigration status. .\n\n###\n\n","completion":" no entity\n , END"}  
{"prompt":"8+ years of development experience with Python and Go or C++ Experience with SDN (e.g. OpenFlow, VPP, Open vSwitch) A proven track record of performing on a winning team with a demonstration of bringing products and services to market. A proven track record of system software development. A proven track record of developing software for extremely large scale, distributed environments. A deep understanding of complex system integration and methodologies for proving the correct operation under nominal and faulted conditions of a hyperscale cloud environment. A deep understanding of developing for high availability and its inherent problems. A deep understanding of containers and container clusters such as kubernetes. A deep understanding of virtual machine environment including VMware, KVM, or Xen. Understanding of Open Systems Interconnection Reference Model (OSI-RM), the TCP\/IP stack, and development experience in L2, L3, L4. Preferred Tech and Prof Experience Familiarity with networking protocols which might include: BGP, DHCP, DNS, OSPF, IGMP, IPv4\/IPv6, IS-IS, and others. Familiarity with Network function virtualization (NFV) functionality and network encapsulation including L4 Load Balancers, Firewalls, etc. Working knowledge of git and CICD tools such as Jenkins and Zuul. EO Statement IBM is committed to creating a diverse environment and is proud to be an equal opportunity employer. All qualified applicants will receive consideration for employment without regard to race, color, religion, gender, gender identity or expression, sexual orientation, national origin, genetics, disability, age, or veteran status. IBM is also committed to compliance with all fair employment practices regarding citizenship and immigration status. .\n\n###\n\n","completion":" experience:['8 + years']\nskills:['python', 'c++']\n END"}  
{"prompt":"Object-Oriented development and design principles. Great communication skills. DevOps Skills Preferred Tech and Prof Experience Experience with Golang. Experience with large scale distributed systems. EO Statement IBM is committed to creating a diverse environment and is proud to be an equal opportunity employer. All qualified applicants will receive consideration for employment without regard to race, color, religion, gender, gender identity or expression, sexual orientation, national origin, genetics, disability, age, or veteran status. IBM is also committed to compliance with all fair employment practices regarding citizenship and immigration status. .\n\n###\n\n","completion":" no entity\n , END"}  
{"prompt":"Basic : - 3+ years of software engineering with an emphasis on system level automation. - 3+ year of experience with a service running in the Microsoft Cloud or similar medium-scale services. - Great oral and written communication skills Preferred : - 5+ years of software industry experience running a service, preferably in the consumer Internet space, with service reliability exceeding 99.9%. - 5+ years of experience managing and engineering solutions using SQL Server or Azure SQL at scale and in Highly Available (HA) environments, with a preferred focus on performance and capacity analysis. - Understanding of large-scale online service network architectures including load balancing, GTM, ACLs, routing, network captures, etc. - Strong working knowledge with Windows 2016, and IIS, including Active Directory, TCP\/IP protocols, and security hardening procedures. - Experience in data structures, algorithms and complexity analysis. - Experience with Microsoft internal tools and processes, experience with deployment technologies (Autopilot, Azure) is a plus. - BA\/BS\/MS degree in CS, related discipline or equivalent work experience.\n\n###\n\n","completion":" experience:['3 + years', '3 + year', '5 + years', '5 + years']\nskills:['software engineering', 'system level automation', 'microsoft cloud', 'software industry', 'consumer internet space', 'managing', 'engineering solutions', 'sql server', 'azure sql', 'highly available ( ha )']\ndiploma:['ba \/ bs \/ ms']\ndiploma\_major:['cs']\n END"}  
{"prompt":"Basic : A BS\/MS in Computer Science or related field Preferred : 2+ years of programming experience writing code in Java, C++, C#, or C or other object-oriented programming language Experience developing and testing computer software and\/or online services Strong coding, debugging and problem-solving skills Strong knowledge of object-oriented programming language paradigms Great communication skills to collaborate cross-group and work effectively within the team Expertise in web or mobile application development Expertise in relational databases, distributed systems and\/or big data technologies Experience in developing large scale services Ability to meet Microsoft, customer and\/or government security screening requirements are required for this role. These requirements include, but are not limited to the following specialized security screenings: Microsoft Cloud Background Check: This position will be required to pass the Microsoft Cloud background check upon hire\/transfer and every two years thereafter.\n\n###\n\n","completion":" diploma:['bs \/ ms']\ndiploma\_major:['computer science']\nexperience:['2 + years']\nskills:['programming experience', 'java', 'c++', 'c #', 'c']\n END"}  
{"prompt":"BS in Electrical Engineering or Computer Engineering, or related Engineering Degree, with 3+ years of industry experience. 3+ years of experience with PCB design with both integrated and discrete circuits, and high-speed interfaces. Experience with development and bring-up of data center system equipment, or consumer devices. Experience with analyzing and resolving system hardware issues. PREFERRED MS in electrical engineering or computer engineering. 6+ years of industry experience. Familiar with server system architecture of one of major CPU suppliers. Experience with end-to-end enabling, design, and deployment cycle of a product. Experience with owning a complete hardware system as hardware lead. Experience with working with thermal, mechanical, management firmware, system firmware, and software teams during product development. Experience with design and troubleshooting data buses used in server such as PCIe, SAS, SATA, I2C. Experience with industry standard EDA tools for developing PCB designs. Knowledge of using Linux for testing.\n\n###\n\n","completion":" diploma:['bs', 'ms']\ndiploma\_major:['electrical engineering', 'computer engineering', 'electrical engineering', 'computer engineering']\nexperience:['3 + years', '6 + years']\nskills:['pcb design', 'integrated', 'discrete circuits', 'high - speed interfaces', 'industry experience']\n END"}  
{"prompt":"Basic : BS\/BA in Electrical\/Mechanical Engineering, Computer Science, Math, Telecommunications, or equivalent experience 4+ Years experience managing large-scale and complex projects 4+ years experience leading technical project management teams with a technical background in network engineering or optical transport Preferred Skills and Experiences Strong working knowledge of physical IT and CE infrastructures (Enterprise or larger business context) Excellent business acumen and financial management experience Experience with Photonics and fiber optics Demonstrated experience leading a team Background Check Requirements: Candidates must meet Microsoft, customer and government security background screening requirements for this role. These requirements currently include, but may change, and are not limited to: Microsoft Cloud Screen Background Check: This position will be required to pass the Microsoft Cloud Screen background check upon hire\/transfer and every two years after that while in this role\n\n###\n\n","completion":" diploma:['bs \/ ba']\ndiploma\_major:['electrical \/ mechanical engineering', 'computer science', 'math', 'telecommunications']\nexperience:['4 + years', '4 + years']\nskills:['managing large - scale', 'complex projects', 'leading technical project']\n END"}

The next step is to setup the OpenAI API. To start, we must install the Python SDK and set our API key using the following commands:

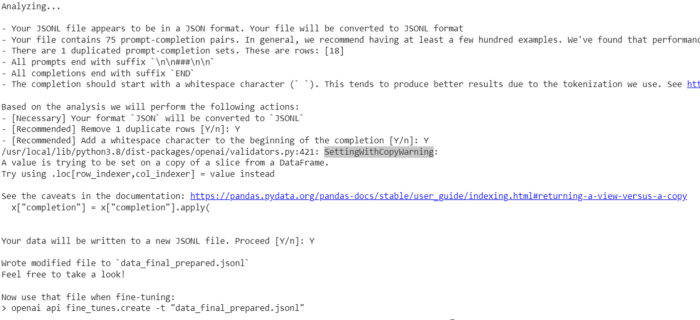


Install OpenAI SDK

We can then use the installed command-line tool to verify our data file with the following command:

https://miro.medium.com/max/700/1*koN2WU4LuACs6tJB_QcPtg.png

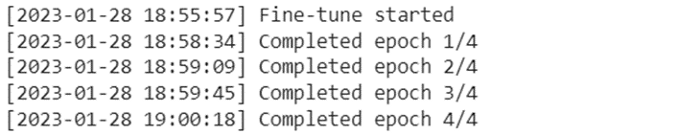
This tool provides recommendations for formatting the data for optimal finetuning if necessary:



Data Formatting Check

We are finally ready to perform the actual finetuning with the following command:

!openai api fine\_tunes.create -t data\_final\_prepared.jsonl -m davinci



Fine-tuning in Progress

The OpenAI API provides a range of base GPT-3 models, among which the Davinci series stands out as the most powerful and advanced, albeit with the highest usage cost. In contrast, the Ada series is the smallest and lightest base model. Although the model sizes are not officially disclosed by OpenAI, it is believed that the Ada, Babbage, Curie, and Davinci models align roughly with 350 million, 1.3 billion, 6.7 billion, and 175 billion parameters, respectively, based on information from their research paper.

For this tutorial, the Davinci model was selected as the optimal choice. Once the fine-tuning process is initiated, the duration can range from 5 to 10 minutes, depending on several factors such as API traffic and the number of examples provided for fine-tuning. Fine-tuning with a training dataset of 75 examples took approximately 5 to 7 minutes. It is worth noting that the duration of the fine-tuning process can vary substantially, depending on the particular scenario and use case.

**Named Entity Extraction with GPT-3**

To test our fine-tuned model, we are going to run it on the job description below:

prompt="""A BS/MS in Computer Science or related field Preferred : 2+ years of programming experience writing code in Java, C++, C#, or C or other object-oriented programming language Experience developing and testing computer software and/or online services Strong coding, debugging and problem-solving skills Strong knowledge of object-oriented programming language paradigms Great communication skills to collaborate cross-group and work effectively within the team Expertise in web or mobile application development Expertise in relational databases, distributed systems and/or big data technologies Experience in developing large scale services Ability to meet Microsoft, customer and/or government security screening requirements are required for this role. These requirements include, but are not limited to the following specialized security screenings: Microsoft Cloud Background Check: This position will be required to pass the Microsoft Cloud background check upon hire/transfer and every two years thereafter. \n\n###\n\n"""

Let’s check the results:



NER using fine-tuned GPT-3 Model

Overall, this model demonstrates effective entity recognition capabilities despite its training on a relatively limited dataset of 75 examples.

**Conclusion**

In conclusion, preparing high-quality training data is essential steps in fine-tuning GPT-3 for specific use cases. With the right training data, GPT-3 can be customized to suit a wide range of applications and produce excellent results.

Moving forward, future work could include expanding the task to relation extraction and document classification, experimenting with different base models to identify the best option for a particular use case, and exploring ways to optimize the fine-tuning process for efficiency and accuracy. If you’re looking to create high quality training data to fine-tune GPT-3 models, the [UBIAI Annotation Tool](https://ubiai.tools/) is definitely worth exploring. With its model-assisted labeling capabilities, seamless collaboration features, and support for various types of documents, you’ll have everything you need to streamline the data annotation process and create high-quality training data in no time.