# **Contact information for Official Representative:**

**Name:** Pradeep Jain

**Email:** Pradeep.Jain@ictect.com

**Team Name:** EUL.ai

# **Names of additional team members:**

**Name:** Prashant Bhoyar

**Name:** Jonathan Boyd

**Name:** Varun Singh

**Name:** Flora Wang

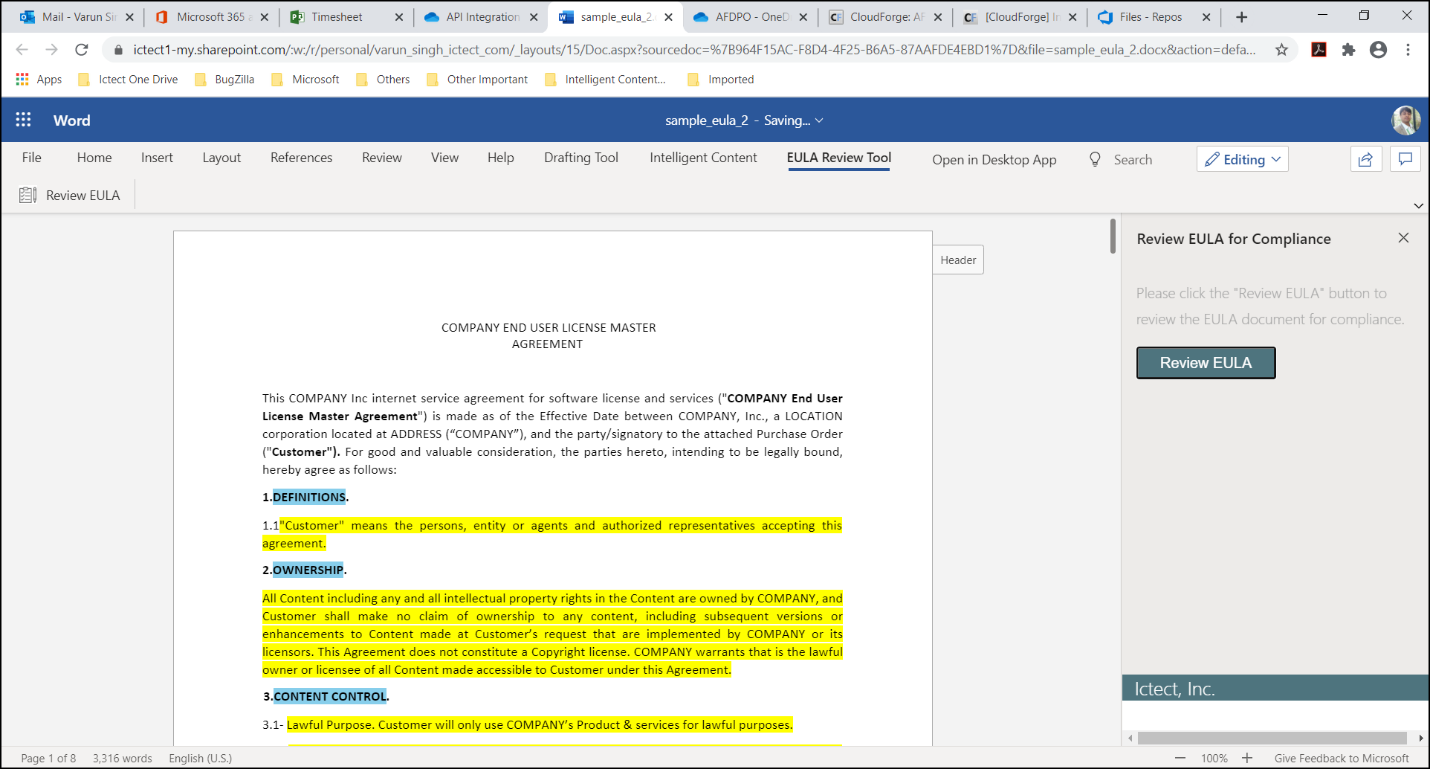
**Name:** Dave Zvenyach

# **Introduction to Team:**

EUL.ai specializes in applying Intelligent Content and Artificial Intelligence in real-world situations. Our team members come from [Ictect](http://www.ictect.com/), [Hangar](https://www.hangar.is/) and include a Microsoft [Most Valuable Professional (MVP)](https://mvp.microsoft.com/en-us/Overview) in Artificial Intelligence. Our companies have experience in bringing innovation in federal agencies.

# **Executive Summary of Solution:**

Our solution comes in the form of a plug-in to Microsoft Word (desktop-based) or Office 365 (cloud-based). It is an aid to the Contracting Officers (CO) in reviewing compliance in EULA documents. The solution appears as a new tab in Word called “EULA Review Tool”. The user opens the EULA document and clicks on a button “Review EULA”. The software reviews each clause, by parsing content and sending it to the AI server. For each clause, the software shows compliance/non-compliance. Please see below a screenshot.



Notes:

1. While the current solution is in the form of a plug-in to Microsoft Word (desktop-based) or Office 365 (cloud-based), we could integrate it to other GSA systems or make it available in other forms. The reason we have used Microsoft Word is because Contracting Officers and users are familiar with it.

2. The current prototype version opens Word documents. PDF can be opened in Word as well. In the future, we plan to provide additional support and ease of use in reviewing PDF documents.

3. The current prototype does not show the reasons for identifying clauses as compliant/non-compliant. We plan to add this reasoning in a future version. We plan to show this in the form of Comments in Microsoft Word/Office 365.

4. Please note that the commercial version of Intelligent Content software, licensed by Ictect, is not used in the prototype. The commercial version enables “rule-based compliance review”. We plan to add it in the future versions.

# **EULA Review Solution Architecture:**

## **Technology Scope:**

* Microsoft Office 365
* Intelligent Content software from Ictect (incorporated in commercial version)
* Support Vector Machines (SVM) and Decision Tree models for AI/ML. Additional models can be added in future versions.
* The prototype uses [Heroku](https://www.heroku.com/) platform. Future versions can be integrated with GSA’s chosen platform.

## **Functionality and User Interface:**

* The software provide interface within Microsoft Word or Office 365. It can be used on the desktop or cloud.
* The current prototype supports MS Word EULA files. PDF can be opened as well, using desktop version of Word. However, we have not conducted any tests using PDF. Future versions will fully support PDF documents from the solution.
* Current version is an interactive aid tool for users. It shows potentially non-compliant clauses for them to review. We can provide batch functionality for reviews in a future version.

## **Application of Artificial Intelligence/Machine Learning (AI/ML):**

* We used a supervised learning approach since GSA has provided a pre-classified set of clauses labeled as valid or invalid.
* After setting aside our own validation set to run metrics on, we trained various types of machine learning models, including a support vector machine, logistic regression, random forest, decision tree, and boosted trees. We chose these methods as opposed to a type of neural network since these methods are able to run faster. We are also able to better explain the results and gather the most important features (words) that contribute to the model’s predictive power. We found neural networks to be unnecessarily complicated and a “black box” approach for this task.
* After running the metrics on our trained models, we decided on a random forest model, which uses an ensemble of decision trees, to determine the predicted probabilities of classifying clauses as valid or invalid.
* From the predicted probabilities, we calculated the optimal threshold for the ultimate classification from the true positive and false positive rates that lie on an ROC curve.

NOTE: Please do not submit any sensitive or classified information.