CMP2804M Team Software Engineering
Team Number: 7
Robot Waiter

Amanpreet, Ben Purchase, Jozef Bonnar, Phuong Doan, Thomas Bowles

University of Lincoln, School of Engineering and Physical Sciences
27691041@students.lincoln.ac.uk, 27619849@students.lincoln.ac.uk, 27497032@students.lincoln.ac.uk, 28101897@students.lincoln.ac.uk, 27659227@students.lincoln.ac.uk

1 Project Proposal

This project involves developing a software solution for a robot waiter, indented for use in a bar or restaurant setting. The main function of this robot is to be in a previous unknown environment and be able to identify customers who wave at the robot signalling that they need to have their order taken, the robot will then navigate to the customer, take their order, and then return to its starting point to communicate the order back to the bartender. The majority of this project will be conducted in simulation, but it may also be tested in real-world scenarios using the TIAGo robot (PAL Robotics, 2024). The robot's modules will focus on integrating various components together so that they work in harmony with each other. These include gesture recognition, navigation and speech recognition to be able to interact with the customer.

1.1 Aims and Objectives

We aim to develop a software solution for a robot waiter at a bar and/or restaurant. The robot will be asked to identify people in a room if waiving and approach them. Once there, the robot needs to record the order and go back to the original location to communicate with a bartender. Our project could be divided into four main components:

- **Develop a program that can identity waving customers in an unstructured environment**
  The goal of this objective is to create a computer vision system that uses the RGB camera to detect and identify waving gestures from customers in an unstructured and changing environment. The application will need to recognise the customer calls to allow the robot to then navigate to the customer. It will need to be optimised for varying lighting conditions and crowded environments.
- **Develop a program that can listen to user input and successfully take a restaurant order**
  Develop a program that can recognise speech to capture and process customers' orders. The goal of the program is to convert speech to text with high accuracy, even in noisy environments such as is found in busy areas like restaurants or cafes. The program will then converse with the customers using natural conversation, enabling further refining of the order by double checking with the customers and clarifying anything necessary. Finally, the program will process the gathered information into a standardised written order for passing on to the staff.
- **Develop a program for robot navigation and path planning in a previously unknown environment**
  The goal of this objective is to develop a 3D mapping solution using point cloud data coming from the 3D Lidar sensors of the robot. This system enables the robot to navigate in an unknown, unstructured, and large environment, where traditional 2D mapping is difficult due to the complexity of the scenario. The program will address obstacles such as moving furniture, customers and more with accurate path planning.

- **Integrate software components into a complete program**
  For this objective we are aiming to integrate the software solution for the three divided tasks – gesture recognition, speech recognition, and 3D mapping – into a unified system. This integration will be our complete system for a robot waiter ensuring the robot's seamless operation from detecting customers, reaching them, taking orders, and transferring orders to a human staff.

1.2 Review

- Gesture recognition has been used in many applications in robotics and human interactions systems. Frameworks like OpenPose (CMU-Perceptual-Computing-Lab, 2020) have been widely adopted for real time multi person pose estimation, extracting skeletal data with high precision (Cao et al., 2017). Its ability to handle occlusions and varying lighting conditions makes it suitable for dynamic environments like restaurants (Sengar et al., 2024). It also has been used for recognising hand gestures in healthcare robots (Fiorini et al., 2021). Another prominent approach is YOLO (You Only Look Once), a high-speed object detection framework often adapted for gesture detection (Redmon et al., 2015). It has proven to be effective as a YOLO-based gesture recognition system for autonomous vehicles demonstrated a robust detection even in cluttered scenes (Sarda et al., 2021). These methods align with our project's goal to create a gesture recognition system that achieves at least a 70% accuracy in detecting waving gestures, crucial for initiating customer interaction in a busy restaurant.
- Speech recognition techniques have been used in various projects throughout the world including navigation, voice assistants, etc. People interact with speech recognition systems all the time in their day to day lives using devices such as Amazon's 'Alexa' (Amazon, 2010) or Apple's Siri (Apple, 2023). These software show how effective natural language processing can be in accurately interpreting user commands. Another similar project is the Pepper robot (Softbank Robotics, 2024), which uses speech recognition to interact with customers in various scenarios, showcasing its potential in roles for use in a restaurant. However, Pepper primarily relies on pre-programmed responses and limited conversational depth, which can restrict its adaptability in dynamic environments like our one. Our project will build on these fallbacks, by integrating Python libraries such as SpeechRecognition (Python Software Foundation, 2024) and pyttsx3 (Python Software Foundation, n.d.) with advanced large language modules like GPT-3.5 (OpenAI, 2024), and Llama (Meta, 2022). These models show a potential to integrate natural language into our robot. Furthermore, we could also leverage the LLM's ability to 'comprehend' the intricacies of human speech in order to understand and list the different variables of an order more accurately. For example, in a restaurant scenario where "Carbonara" is listed on the menu, if the customer asks for "pasta," the LLM can use its interpretative abilities to clarify whether the customer intended to order the "Carbonara."
- Navigation and path planning is a challenge for any robot, one company who is one of the leaders in this space is Boston Dynamics (Boston Dynamics, 2024). Their robots, such as Spot (Boston Dynamics, 2023) demonstrate exceptional capabilities in navigating complex and dynamic terrains from warehouses, launchpads and much more in real time. Spot utilises a combination of sensors, including Lidar and depth cameras, alongside a robust path planning algorithm. This allows it to handle any unexpected object, allowing it to adapt to various environments, almost making it a benchmark for autonomous navigation in robotics. Our project will use similar principles. We will be using the ROS (ROS Community, undated*a*) and OctoMap (OctoMap, 2022) based navigation systems to integrate 3D LIDAR data into real time maps. From there we will then flatten them into 2D occupancy grids. We will try to optimise the path planning for restaurant environments by addressing challenges like the shifting of chairs, people and more. Simulation tools like Gazebo (Robotics, undated) and Rviz (ROS Community, undated*b*) will allow us to refine these capabilities, ensuring responsiveness, accuracy and safety in a controlled environment before deployment.
- Multi-modal foundation models have been actively developed to enable robots to perform complex tasks in unstructured environments, combining visual data with linguistic cues to interpret, plan, and execute actions. A system utilising the augmented reality-gesture-voice integration in human-robot interaction enhancing navigation, object detection, and diver's wave has been proposed for underwater exploration and rescue missions (Valluri, 2024). In social contexts, a model integrating vision and language processing system to infer emotions of caregivers has shown a potential in accuracy enhancement (Lee et al., 2024). Regrading the system integration goal in our project, we aim to build upon these advancements with a focus on achieving seamless interaction of robot with customers/restaurant staffs through its understanding of the environment, human's gesture, and natural language.

1.3 Schedule and Risks

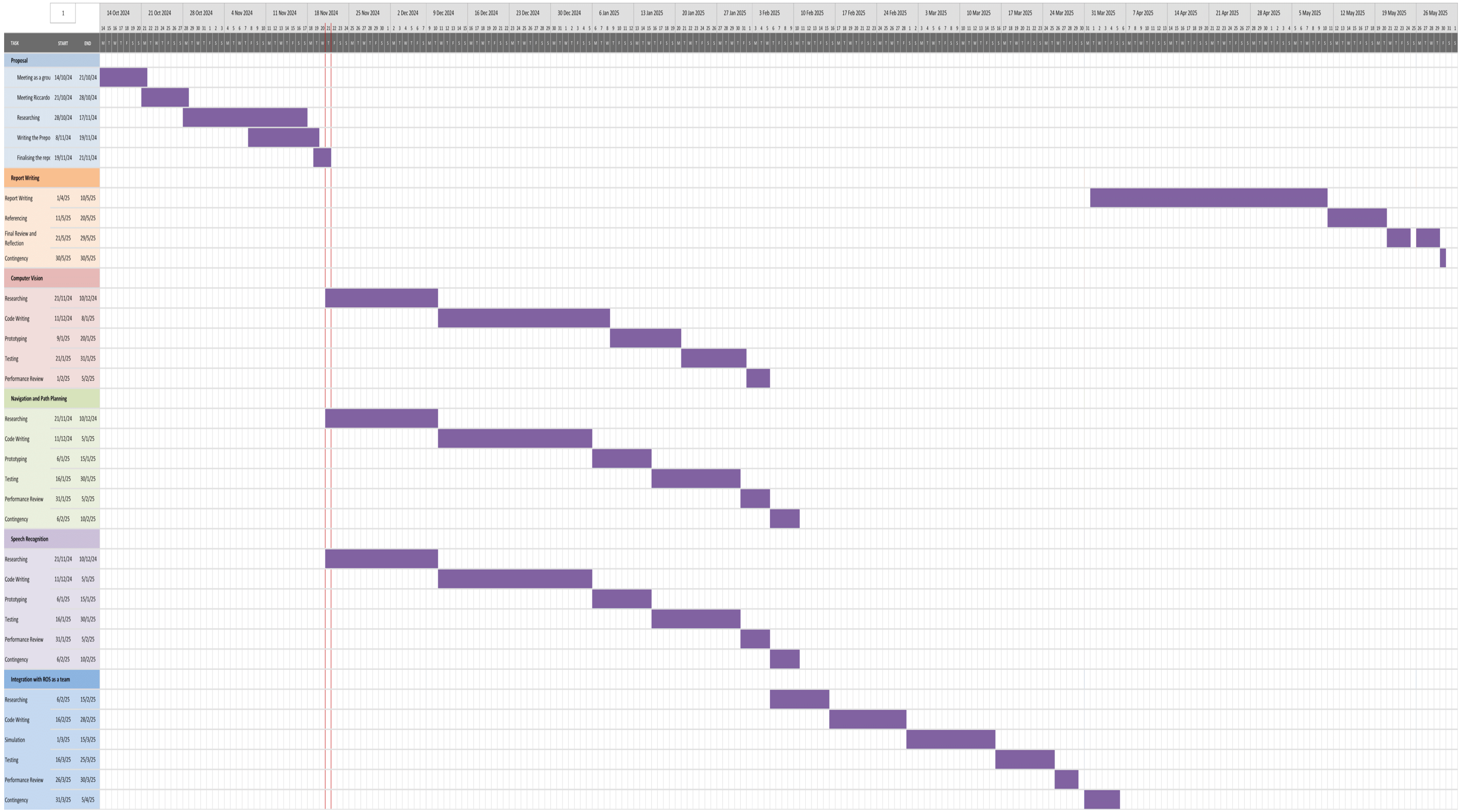

Fig. 1. Project's Gantt chart

Risk	**Severity**	**Mitigation**
Team communication breakdowns	High	Schedule regular meetings, use collaborative tools (shared documents, Discord), and establish clear communication.
Uneven contribution among team members	Medium	Set clear roles and responsibilities, maintain a shared task tracker, and conduct periodic team reviews to ensure accountability.
Misaligned timelines or project delays	High	Create a detailed project timeline with buffer periods for unforeseen delays and hold weekly progress checkpoints.
Conflicts within the team	Medium	Establish a conflict resolution framework provided in the module document, encourage open communication, and involve a mediator if necessary.
Loss of group cohesion due to remote working	High	Use video conferencing tools for meetings, maintain regular updates, and encourage informal check-ins to strengthen cohesion.
Loss of a team member completely from the group project	High	Redistribute the tasks among the team, review the deadline and ensure they are met, and speak with the supervisor.

Table 1. Risk Assessment and Mitigation Strategies

2 Legal, Social, Ethical and Professional Issues

Throughout planning the project, we realised there are many blaring LSEPI regarding robotics. The privacy of personal data stands out as the first thing to consider; to comply with the Data Protection Act 1998 (Legislation, 1998) and the GDPR (Intersoft Consulting, 2016), we will delete stored visual and audio data captured by the robot of the customer upon it being no longer needed for the robot to carry out its tasks. This inhibits the possibility of us using/distributing personal data without a customer's consent. We will also make it clear that the robot will store personal data for the shortest period of time to complete their order (acquiring verbal consent).

We will ensure that gender and racial bias is minimised as much as we possibly can to ensure equality of treatment regarding customers is upheld without fail regardless of gender or race as per the Equality Act 2010 (Legislation, 2010). Eliminating bias is crucial to maintaining the wellbeing of customers who will interact with the robot which is a fundamental to maintaining a good ethical standing.

Another thing to consider is the ethical issues of replacing humans in the workplace (job displacement). Whilst it can be agreed that losing humanity in social interactions can have negative effects on the social skills of customers, we believe that this is a non-issue as it is extremely unlikely that a complete robot takeover of the workplace is plausible anytime soon nor would we push that agenda. The use of robots in a predominantly human workspace could also contribute to economic inequality between companies and workers. We would opt to require the robot to be deployed as a single unit to minimise loss of humanity in social environments and limit economic inequality. The contribution to work shortages is also clear as we will inevitably be replacing the need of a human waitress/waiter with a robot. But we will make certain any redundancies are made known to current employees and compensate them with a severance package which would give them the means to look for satisfactory work elsewhere.

In accordance with ethical and legal standards a risk assessment needs to be made as the liability of customers falls on us (the robot programmers). Robots are prone to malfunction and do unexpected things, especially in a restaurant environment with many unknowns. An emergency shutdown needs to be implemented to ensure no harm can happen to customers due to a robot's actions (even unintentional harm is our responsibility to prevent). We recognise the importance of safety and reliability when it comes to robotics and will ensure that the probability of harm being caused is at a negligible amount.

After deploying a robot, maintenance is crucial to upkeep our outlined standards of function. To comply with legal requirements and to maintain our professional standards we must perform regular checks on the robot to remain vigilant on potential issues which may arise after a short time (bi-weekly maintenance checks).

In conclusion, we will work to uphold a professional standard akin to ACM (Association for Computing Machinery, 2018) by respecting customer privacy, avoiding harm, providing a risk assessment and mitigating discrimination of customers whilst also maintaining full accountability of all actions the robot performs regardless of whether it was our intention or not.

3 Group Work Reflection

Team reflection:

- **Strength:** we have consistently organised and attended weekly meetings allowing for strong communication and with our supervisor allowed for project resources, direction and feedback. From there, specific task allocation were done early on, which is beneficial for a better project management and risk assessment. Communication within the group, between us and our supervisor has been effectively managed by selecting a team leader working as a liaison.
- **Limitations:** As we did not thoroughly planned and produced a gantt chart in our first step, both the team and our supervisor were put in a rush to finish our proposal on time, resulting in lack of time for a more careful refinement and feedback from our supervisor. We should have set more stricter deadlines with more progression reviews and divided the proposal writing tasks more specifically. This experience will be reflected for improvements in our next stages.
- **Challenges:** Having a team member being placed in a wrong level grade at the beginning resulted in time wasting trying to contact him. Also, we had some availability conflicts as sometimes not all members were always available for meetings, and to address this problem, our team policies were formed regarding informed absences and catch up. In semester B, a team member's exchange term will additionally present new challenges in coordinating meetings and tasks, requiring us to think of adjustments like virtual meetings. Finally, our team all has limited prior experience and knowledge with robotics and the project domain, which required extra time for researching and learning.

Individual reflection:

1. **Amanpreet** has focused on researching risks in robot navigation task, contributed to outlining policies and procedures for our team contract. Lack of clarity in certain areas of the proposal was a limitation preventing him from grasping the scope of my responsibilities.
2. **Ben Purchase** has put his effort into researching the speech recognition task. The wide variance in accessibility proved to be a limitation, and his busy schedule requires a better time management.
3. **Jozef Bonnar** has been focusing on the gesture recognition and integration with ROS, he is the most familiar with the GitHub and simulation provided due to experience in the past. An improvement would be trying to meet deadlines with time to spare instead of leaving it till the day before the meeting.
4. **Phuong Doan** has been working as liaison effectively ensuring the our connection with our supervisor with good meeting schedules and management. More urgency in completing work could have improved her section of proposal for more peer review and adjustments.

5. **Thomas Bowles** has conducted thorough research into LSEPI issues regarding robotics to complete his assigned task to an appropriate standard. More urgency in completing work could have improved his section of the proposal for more peer review and adjustments.

4 Table of Contributions

Table 2. *Team member contribution percentages.*

Team Member	**Contribution Level**
Amanpreet	Contribution 100%
Ben Purchase	Contribution 100%
Jozef Bonnar	Contribution 100%
Phuong Doan	Contribution 100%
Thomas Bowles	Contribution 100%

REFERENCES

Amazon (2010), ‘Alexa’. [accessed: 17 November 2024].
URL: *https://developer.amazon.com/en-US/alexa*

Apple (2023), ‘Siri’. [accessed: 17 November 2024].
URL: *https://www.apple.com/siri/*

Association for Computing Machinery (2018), ‘Acm code of ethics and professional conduct’. [accessed: 19 November 2024].
URL: *https://www.acm.org/code-of-ethics*

Boston Dynamics (2023), ‘Spot – the agile mobile robot’. [accessed: 15 November 2024].
URL: *https://bostondynamics.com/products/spot*

Boston Dynamics (2024), ‘Home’. [accessed: 15 November 2024].
URL: *https://bostondynamics.com/*

Cao, Z., Simon, T., Wei, S.-E. and Sheikh, Y. (2017), Realtime multi-person 2d pose estimation using part affinity fields, *in* ‘IEEE Conference on Computer Vision and Pattern Recognition (CVPR)’. [accessed: 12 November 2024].
URL: *https://doi.org/10.1109/cvpr.2017.143*

CMU-Perceptual-Computing-Lab (2020), ‘Openpose: Real-time multi-person keypoint detection library for body, face, hands, and foot estimation’. [accessed: 20 November 2024].
URL: *https://github.com/CMU-Perceptual-Computing-Lab/openpose*

Fiorini, L., Loizzo, F., Sorrentino, A., Kim, J., Rovini, E., Di Nuovo, A. and Cavallo, F. (2021), ‘Daily gesture recognition during human-robot interaction combining vision and wearable systems’, *IEEE Sensors Journal* **21**(20), 23568–23577. [accessed: 10 November 2024].
URL: *https://doi.org/10.1109/jsen.2021.3108011*

Intersoft Consulting (2016), ‘General data protection regulation (gdpr)’. [accessed: 19 November 2024].
URL: *https://gdpr-info.eu/*

Lee, S., Lee, S. and Park, H. (2024), ‘Integration of tracking, re-identification, and gesture recognition for facilitating human-robot interaction’, *Sensors* **24**(15), 48–50. [accessed: 19 November 2024].
URL: *https://www.mdpi.com/1424-8220/24/15/4850*

Legislation (1998), ‘Data protection act 1998’. [accessed: 20 November 2024].
URL: *https://www.legislation.gov.uk/ukpga/1998/29/contents*

Legislation (2010), ‘Equality act 2010’. [accessed: 21 November 2024].
URL: *https://www.legislation.gov.uk/ukpga/2010/15/contents*

Meta (2022), ‘Llama 3.2’. [accessed: 9 November 2024].
URL: *https://www.llama.com/*

OctoMap (2022), ‘Octomap - an efficient probabilistic 3d mapping framework based on octrees’. [accessed: 20 November 2024].
URL: *https://github.com/OctoMap/octomap*

OpenAI (2024), ‘Openai documentation’. [accessed: 20 November 2024].
URL: *https://platform.openai.com/docs/models*

PAL Robotics (2024), ‘Tiago - mobile manipulator robot’. [accessed: 20 November 2024].
URL: *https://pal-robotics.com/robot/tiago/*

Python Software Foundation (2024), ‘Speech recognition 3.11.0’. [accessed: 20 November 2024].
URL: *https://pypi.org/project/SpeechRecognition/*

Python Software Foundation (n.d.), ‘pyttsx3: Text-to-speech library’. [accessed: 20 November 2024].
URL: *https://pypi.org/project/pyttsx3/*

Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2015), ‘You only look once: Unified, real-time object detection’, *arXiv preprint* . [accessed: 10 November 2024].
URL: *https://arxiv.org/abs/1506.02640*

Robotics, O. (undated), ‘Gazebo simulator’. [accessed: 19 November 2024].
URL: *https://gazebosim.org/home*

ROS Community (undated*a*), ‘Robot operating system (ros)’. [accessed: 21 November 2024].
URL: *https://www.ros.org/*

ROS Community (undated*b*), ‘Rviz - ros wiki’. [accessed: 21 November 2024].
URL: *https://wiki.ros.org/rviz*

Sarda, A., Dixit, S. and Bhan, A. (2021), Object detection for autonomous driving using yolo [you only look once] algorithm, *in* 'IEEE International Conference on Computing, Communication, and Vision (ICICV)'. [accessed: 2 November 2024].
URL: *https://doi.org/10.1109/ICICV50876.2021.9388577*

Sengar, S., Kumar, A. and Singh, O. (2024), 'Efficient human pose estimation: Leveraging advanced techniques with mediapipe', *arXiv preprint* . [accessed: 19 November 2024].
URL: *https://arxiv.org/html/2406.15649v1*

Softbank Robotics (2024), 'Meet pepper'. [accessed: 20 November 2024].
URL: *https://us.softbankrobotics.com/pepper*

Valluri, D. (2024), 'Integrating augmented reality, gesture recognition, and nlp for enhancing underwater human-robot interaction', *International Journal of Science and Research Archive* **11**(02), 956–968. [accessed: 19 November 2024].
URL: *https://ijsra.net/sites/default/files/IJSRA-2024-0509.pdf*

A Team Contract

GOALS

1. Our objective is to achieve a first-class grade by ensuring that every aspect of the assignment meets or exceeds the highest standards, demonstrating excellence in both the content and execution of the project.
2. We are working on building a strong team dynamic by fostering mutual respect, accountability, and support among all members.
3. We are aiming to gain knowledge in robotics concepts, including speech & gesture recognition, navigation, mapping, path planning, etc., and robotics tools such as ROS, Lidar and sensors. We aim to apply our theoretical knowledge to real-world scenarios.

EXPECTATIONS

1. We commit to attending all scheduled meetings and will inform the team in advance if unable to attend. Members who miss meetings will stay updated by reviewing notes and contributing asynchronously.
2. Every team member is responsible for completing their assigned tasks to the best of their ability. We pledge to actively engage in discussions, share ideas, and fulfill our roles and responsibilities diligently. This ensures fairness and prevents undue burden on others.
3. We are making sure not to step on each other's toes by respecting each team member's responsibilities and contributions. We collaborate by offering support and suggestions without overstepping boundaries.
4. We respect each other's contributions and provide constructive feedback when needed to ensure continuous improvement and collaboration.

POLICIES & PROCEDURES

1. The team shall provide support for team members' individual concerns.
2. We commit to using the specified channels of communication (Discord) regularly and efficiently to keep everyone informed. To preserve workflow and prevent needless delays, messages and requests should be answered within 24 hours.
3. Meetings as a team are necessary for cooperation and advancement. Attendance and full participation in discussions and decision-making are required of every member. Meeting productivity and efficiency are largely dependent on preparedness and punctuality.
4. Meeting absence needs to be informed at least 24 hours before the meeting. The team shall provide meeting notes and updates for approved absence.
5. We accept accountability for finishing the tasks we are given on schedule and to the best possible quality. It's critical to notify the team as soon as any challenges appear so that we can modify our plans and move forward.
6. The team commits to producing a consistent quality, well-commented code, working to a high expectation, and providing detailed reporting.

CONSEQUENCES

To address non-performance in regard to our proposed goals, expectations, policies and procedures, the team agree to follow instructions in the Dispute Resolution Document provided on the module site.

1. Members of the team will be reminded that non-performance can impact not only the individual but also the group's capacity to accomplish deadlines and objectives.
2. If a team member is not meeting expectations or fulfilling responsibilities, a private discussion will first be held within the team to understand the issue, offer support, and agree on a plan to improve.
3. If performance issues persist, the team will have a follow-up conversation to discuss further steps and potential adjustments to responsibilities.
4. If the non-performance then still persists, the matter shall be brought to the attention of the project supervisor (Dr. Riccardo Polvara) for additional resolution.
5.

Table 3. Team Contract

We share these goals and expectations, and agree to these policies, procedures, and consequences.

Amanpreet (signed)

Ben Purchase (signed)

Jozef Bonnar (signed)

Phuong Doan (signed)

Thomas Bowles (signed)
