



THE PROGRAM GUIDE FOR
Fifth AAAI/ACM Conference on
ARTIFICIAL
INTELLIGENCE,
ETHICS,
& SOCIETY

AUGUST 1-3, 2022 | OXFORD, ENGLAND

 Follow AIES 2022 on Twitter! @AIESConf





AIES-22 WIFI ACCESS

Wireless internet is available in all bedrooms and meeting spaces.

For Wi-fi access please connect to: [Keble](#)
Then use password: [RedBrickOxford](#)

Next, open your web browser to be redirected to the Keble registration website, click the “Begin Registration” button, select the name of the conference you are attending, and follow the prompts on screen.

[The conference password is: AI2022](#)

Note: If you are not redirected to the Keble registration website then just try to browse to any non-https webpage (such as www.it.ox.ac.uk) to force the redirection.

If you are attending the meeting as a day delegate (not staying overnight at Keble) please follow the above instructions, when room number is requested enter **0000**.



TRAVELING TO OXFORD

[How to Get to Oxford](#)

[Information on Oxford Park & Rides](#)

[How to find Keble College](#)

A decorative graphic on the left side of the page features a dark blue silhouette of a city skyline at the bottom, including a prominent domed building. Above the skyline, several colored circles (red, blue, and maroon) are scattered across the white background. The main content is a table of contents on the right side of the page.

PROGRAM CONTENTS

Acknowledgements 4

Program Overview 5

Detailed Program 6

Monday 6

Tuesday 10

Wednesday 14

Sponsors 17

Keble College Maps 18

ACKNOWLEDGEMENTS

The Association for the Advancement of Artificial Intelligence acknowledges and thanks the following individuals for their generous contributions of time and energy to the successful creation and planning of the Fifth Annual AAAI / ACM Conference on Artificial Intelligence, Ethics, and Society.

CONFERENCE CO-CHAIRS

Vincent Conitzer *Duke University*
John Tasioulas *Oxford University*

CONFERENCE PROGRAM CO-CHAIRS

Matthias Scheutz *Tufts University*
Ryan Calo *University of Washington*
Martina Mara *Johannes Kepler University Linz*
Annette Zimmermann *University of York*

STUDENT PROGRAM CHAIRS

Brent Venable *University of West Florida*
Emmanuelle Burton *University of Illinois Chicago*
Theodore Lechterman *University of Oxford*

VIRTUAL EXPERIENCE CHAIR

Wenbin Zhang *Carnegie Mellon University*

PROGRAM OVERVIEW

The AIES-22 program will be held virtually and at Keble College, Oxford University. All timings are in British Summer Time (UTC / GMT +1).

MONDAY, AUGUST 1

9:15–9:30

Welcome

9:30–10:15

Invited Talk 1

10:15–10:30

Q&A

10:30–10:45

Coffee Break

10:45–11:45

Lightning Talks

11:45–12:00

Break

12:00–12:45

Invited Talk 2

12:45–1:00

Q&A

1:00–3:00

Lunch Break /
Online Poster Sessions

3:00–4:00

Lightning Talks

4:00–4:15

Coffee Break

4:15–5:00

Invited Talk 3

5:00–5:15

Q&A

5:15–5:30

Break

5:30–7:30

Reception /
Poster Session

TUESDAY, AUGUST 2

Invited Talk 4

Q&A

Coffee Break

Lightning Talks

Break

Invited Talk 5

Q&A

Lunch Break /
Online Poster Sessions

Lightning Talks

Coffee Break

Invited Talk 6

Q&A

Break

Reception /
Poster Session

WEDNESDAY, AUGUST 3

Student Lightning Talks

Student Lightning Talks

Student Lightning Talks

Student Poster Session

Closing Remarks

Farewell Lunch

Farewell Lunch



9:15–9:30 **Welcome // O'Reilly Theatre**
Program and General Chairs

9:30–10:15 **Invited Talk 1: Inioluwa Deborah Raji, University of California, Berkeley // O'Reilly Theatre**
From Algorithmic Audits to Actual Accountability

10:15–10:30 **Q&A**

10:30–10:45 **Coffee Break // Arco Building**

10:45–11:45 **Lightning Talk Session 1 // O'Reilly Theatre**

American == White in Multimodal Language-and-Image AI
Robert Wolfe and Aylin Caliskan

Ordinary People as Moral Heroes and Foes:
Digital Role Model Narratives Propagate Social Norms in China's Social Credit System
Mo Chen, Severin Engelmann and Jens Grossklags

A Meta-Analysis of the Utility of Explainable Artificial Intelligence in Human-AI Decision-Making
Max Schemmer, Patrick Hemmer, Maximilian Nitsche, Niklas Kühl and Michael Vössing

Does AI de-Bias Recruitment? Race, Gender, and AI's 'Eradication of Differences Between Groups'
Kerry Anne Mackereth and Eleanor Drage

Gender Bias in Word Embeddings: A Comprehensive Analysis of Frequency, Syntax, and Semantics
Aylin Caliskan, Pimparkar Parth Ajay, Tessa Charlesworth, Robert Wolfe and Mahzarin Banaji

Managing Sustainability Tensions in Artificial Intelligence: Insights from Paradox Theory
Eleanor Mill, Wolfgang Garn and Nick Ryman-Tubb

Crowdsourcing Impacts: Exploring the Utility of Crowds for Anticipating Societal
Impacts of Algorithmic Decision Making
Julia Barnett and Nicholas Diakopoulos

Social media profiling continues to partake in the development of formalistic self-concepts.
Social media users think so, too.
Severin Engelmann, Valentin Scheibe, Fiorella Battaglia and Jens Grossklags

FINS Auditing Framework: Group Fairness for Subset Selections
Kathleen Cachel and Elke Rundensteiner

MONDAY, AUGUST 1

Write It Like You See It: Detectable Differences in Clinical Notes By Race Lead To Differential Model Recommendations

Hammaad Adam, Ming Ying Yang, Kenrick Cato, Ioana Baldini, Charles Senteio, Leo Celi, Jiaming Zeng, Moninder Singh and Marzyeh Ghassemi

Artificial Moral Advisors: A New Perspective from Moral Psychology
Yuxin Liu, Adam Moore, Jamie Webb and Shannon Vallor

A Model for Governing Information Sharing in Smart Assistants
Xiao Zhan, Stefan Sarkadi, Natalia Criado Pacheco and Jose Such

11:45–12:00 **Short Break**

12:00–12:45 **Invited Talk 2: Ronald Arkin, Georgia Institute of Technology // O'Reilly Theatre**
Robots that Need to Mislead: Biologically-inspired Machine Deception

12:45–1:00 **Q&A**

1:00–3:00 **Lunch Break / Online Posters // Dining Hall & Zoom Rooms**

Aegis: An Agent for Multi-party Privacy Preservation
Rim Ben Salem

FairCanary: Rapid Continuous Explainable Fairness
Avijit Ghosh

Making Human-Like Moral Decisions
Nicholas Mattei

A Dynamic Decision-Making Framework Promoting Long-Term Fairness
Bhagyashree Puranik

Understanding Decision Subjects' Fairness Perceptions and Retention in repeated Interactions with AI-Based Decision Systems
Meric Altug Gemalmaz

An Ontology for Fairness Metrics
Jade Franklin

Civil War Twin: Exploring Ethical Challenges in Designing an Educational Face Recognition Application
Manisha Kusuma

Dynamic Fleet Management and Household Feedback for Garbage Collection
Martin Aleksandrov

Practical Skills Demand Forecasting via Representation Learning of Temporal Dynamics
Maysa Macedo

MONDAY, AUGUST 1

No Rage Against the Machines: Threat of Automation Does Not Change Policy Preferences
Baobao Zhang

Mimetic Models: Ethical Implications of AI that Acts Like You
Solon Barocas

Contrastive Counterfactual Fairness in Algorithmic Decision-Making
Ece Mutlu

Ordinary People as Moral Heroes and Foes: Digital Role Model Narratives
Propagate Social Norms in China's Social Credit System
Mo Chen

American==White in Multimodal Language-and-Image AI
Robert Wolfe

Resume Format, LinkedIn URLs and Other Unexpected Influences on AI Personality
Prediction in Hiring: Results of an Audit
*Alene Rhea, Kelsey Markey, Lauren D'Arinzo, Hilke Schellmann, Mona Sloane,
Paul Squires and Julia Stoyanovich*

Towards Robust Off-Policy Evaluation via Human Inputs
Harvineet Singh

3:00–4:00 Lightning Talk Session 2 // O'Reilly Theatre

How Cognitive Biases Affect XAI-assisted Decision-making: A Systematic Review
Astrid Bertrand, Rafik Belloum, James R. Eagan and Winston Maxwell

Racial Disparities in the Enforcement of Marijuana Violations in the US
*Bradley Butcher, Christopher Robinson, Miri Zilka, Riccardo Fogliato,
Carolyn Ashurst and Adrian Weller*

Long-term Dynamics of Fairness Intervention in Connection Recommender Systems
Nil-Jana Akpinar, Cyrus DiCiccio, Preetam Nandy and Kinjal Basu

The Cost of Ethical AI Development for AI Startups
James Bessen, Stephen Michael Impink and Robert Seamans

Practical Skills Demand Forecasting via Representation Learning of Temporal Dynamics
*Maysa M G Macedo, Wyatt Clarke, Eli Lucherini, Tyler Baldwin, Dilermando Queiroz,
Rogerio de Paula and Subhro Das*

Algorithms that "Don't See Color": Measuring Biases in Lookalike and Special Ad Audiences
Piotr Sapiezynski, Avijit Ghosh, Levi Kaplan, Alan Mislove and Aaron Rieke

Fairness in Agreement With European Values: An Interdisciplinary Perspective on AI Regulation
*Alejandra Bringas Colmenarejo, Luca Nannini, Alisa Rieger, Kristen Marie Scott, Xuan Zhao,
Gourab K Patro, Gjergji Kasneci and Katharina Kinder-Kurlanda*

MONDAY, AUGUST 1

Towards a Feminist Metaethics of AI
Anastasia Siapka

How Does Predictive Information Affect Human Ethical Preferences?
Saumik Narayanan, Guanghui Yu, Wei Tang, Chien-Ju Ho and Ming Yin

Explainability's Gain is Optimality's Loss? – How Explanations Mislead
Charles Wan, Rodrigo Belo and Leid Zejnilovic

Counterfactual explanations for prediction and diagnosis in XAI
Xinyue Dai, Mark T. Keane, Laurence Shalloo, Elodie Ruelle and Ruth M.J. Byrne

Contrastive Counterfactual Fairness in Algorithmic Decision-Making
Ece Mutlu, Ozlem Ozmen Garibay and Niloofar Yousefi

4:00–4:15 **Coffee Break // Arco Building**

4:15–5:00 **Invited Talk 3: Shannon Vallor, University of Edinburgh // O'Reilly Theatre**
The AI Mirror: Reclaiming our Humanity in an Age of Machine Thinking

5:00–5:15 **Q&A**

5:15–5:30 **Short Break**

5:30–7:30 **Reception / Poster Session 1 // Arco Building**



~~9:30–10:15~~ ~~Invited Talk 4: Abeba Birhane, Mozilla Foundation // O'Reilly Theatre~~ ~~CANCELLED~~
~~The Limits of Fairness~~

~~10:15–10:30~~ ~~Q&A~~ ~~CANCELLED~~

10:30–10:45 **Coffee Break // Arco Building**

10:45–11:45 **Lightning Talk Session 3 // O'Reilly Theatre**

Respect as a Lens for the Design of AI Systems

William Seymour, Max Van Kleek, Reuben Binns and Dave Murray-Rust

Outsider Oversight: Designing a Third Party Audit Ecosystem for AI Governance

Inioluwa Deborah Raji, Peggy Xu, Colleen Honigsberg and Daniel Ho

A Penalty Default Approach to Preemptive Harm Disclosure and Mitigation for AI Systems

Rui-Jie Yew and Dylan Hadfield-Menell

A Bio-Inspired Framework for Machine Bias Interpretation

Jake Robertson, Catherine Stinson and Ting Hu

Algorithmic Fairness and Structural Injustice: Insights from Feminist Political Philosophy

Atoosa Kasirzadeh

Actual Causality and Responsibility Attribution in Decentralized Partially

Observable Markov Decision Processes

Stelios Triantafyllou, Adish Singla and Goran Radanovic

Human Autonomy in Algorithmic Management

Charlotte Unruh, Charlotte Haid, Fottner Johannes and Tim Büthe

A Survey of the Potential Long-term Impacts of AI

Sam Clarke and Jess Whittlestone

Current and Near-Term AI as a Potential Existential Risk Factor

Benjamin Bucknall and Shiri Dori-Hacohen

Responsible AI systems: Who are the stakeholders?

Advait Deshpande and Helen Sharp

Identifying Bias in Data Using Two-Distribution Hypothesis Tests

William Yik, Limnantes Serafini, Timothy Lindsey and George Montanez

Transparency, Governance and Regulation of Algorithmic Tools Deployed in the
Criminal Justice System: a UK Case Study

Miri Zilka, Holli Sargeant and Adrian Weller

TUESDAY, AUGUST 2

Achievement and Fragility of Long-term Equitability
Andrea Simonetto and Ivano Notarnicola

Artificial Concepts of Artificial Intelligence: Institutional Compliance and Resistance in AI Startups
Amy Winecoff and Elizabeth Watkins

Data-Centric Factors in Algorithmic Fairness
Nianyun Li, Naman Goel and Elliott Ash

11:45–12:00 **Short Break**

12:00–12:45 **Invited Talk 5: Bertram Malle, Brown University // O'Reilly Theatre**
Beyond Fairness and Explanation: Foundations of Trustworthiness of Artificial Agents

12:45–1:00 **Q&A**

1:00–3:00 **Lunch Break / Online Posters // Dining Hall & Zoom Rooms**

A Dynamic Decision-Making Framework Promoting Long-Term Fairness
Bhagyashree Puranik

DIME: Fine-grained Interpretations of Multimodal Models via Disentangled Local Explanations
Yiwei Lyu

Understanding Decision Subjects' Fairness Perceptions and Retention in repeated Interactions with AI-Based Decision Systems
Meric Altug Gemalmaz

An Ontology for Fairness Metrics
Jade Franklin

Piecemeal knowledge acquisition for computational normative reasoning
Ilaria Canavotto

Dynamic Fleet Management and Household Feedback for Garbage Collection
Martin Aleksandrov

Practical Skills Demand Forecasting via Representation Learning of Temporal Dynamics
Maysa Macedo

No Rage Against the Machines: Threat of Automation Does Not Change Policy Preferences
Baobao Zhang

Mimetic Models: Ethical Implications of AI that Acts Like You
Solon Barocas

Contrastive Counterfactual Fairness in Algorithmic Decision-Making
Ece Mutlu

From Coded Bias to Existential Threat: Expert Frames and the Epistemic Politics of AI Governance
Hendrik Schopmans

Developing Artificial Intelligence for Good: Interdisciplinary research collaborations and the making of ethical AI
Shira Zilberstein

Auditing for Diversity using Representative Examples
Vijay Keswani

Resume Format, LinkedIn URLs and Other Unexpected Influences on AI Personality Prediction in Hiring: Results of an Audit
Alene Rhea, Kelsey Markey, Lauren D'Arinzo, Hilke Schellmann, Mona Sloane, Paul Squires and Julia Stoyanovich

What are People Talking about in #BlackLivesMatter and #StopAsianHate? Exploring and Categorizing Twitter Topics Emerged in Online Social Movements through the Latent Dirichlet Allocation Model
Luyao Zhang

Computational Philosophy
Hubert Etienne

Fair, Robust, and Data-Efficient Machine Learning in Healthcare
Harvineet Singh

Diversity by Design. Balancing the Protection and Inclusion of Users on Online Social Platforms
Paula Helm

The danger of citizen domination through algorithmic decision-making in the public sector
Jana Misis

Evaluating AI-based Public Policies with Social Simulation
Jackob Gutmann

3:00–4:00 **Lightning Talk Session 4 // O'Reilly Theatre**

Mimetic Models: Ethical Implications of AI that Acts Like You
Reid McIlroy-Young, Jon Kleinberg, Siddhartha Sen, Solon Barocas and Ashton Anderson

Measuring Gender Bias in Word Embeddings of Gendered Languages Requires Disentangling Grammatical Gender Signals
Shiva Omrani Sabbaghi and Aylin Caliskan

Generating Deontic Obligations From Utility-Maximizing Systems
Colin Shea-Blymyer and Houssam Abbas

Stochastic Policies in Morally Constrained (C-)SSPs
Charles Evans, Claire Benn, Ignacio Ojea Quintana, Pamela Robinson and Sylvie Thiebaut

TUESDAY, AUGUST 2

The worst of both worlds: A comparative analysis of errors in learning from data in psychology and machine learning
Jessica Hullman, Sayash Kapoor, Priyanka Nanayakkara, Andrew Gelman and Arvind Narayanan

SCALES: From Fairness Principles to Constrained Decision-Making
Sreejith Balakrishnan, Jianxin Bi and Harold Soh

From Explanation to Recommendation: Ethical Standards for Algorithmic Recourse
Emily Sullivan and Philippe Verreault-Julien

Investigating Debiasing Effects on Classification and Explainability
Marta Marchiori Manerba and Riccardo Guidotti

Equalizing Credit Opportunity in Algorithms: Aligning Algorithmic Fairness Research with U.S. Fair Lending Regulation
I. Elizabeth Kumar, Keegan Hines and John P. Dickerson

How Algorithms Shape the Distribution of Political Advertising: Case Studies of Facebook, Google, and TikTok
Orestis Papakyriakopoulos, Mihir Kshirsagar, Christelle Tessono and Arvind Narayanan

Examining Responsibility and Deliberation in AI Impact Statements and Ethics Reviews
David Liu, Priyanka Nanayakkara, Sarah Sakha, Grace Abuhamad, Su Lin Blodgett, Nicholas Diakopoulos, Jessica Hullman and Tina Eliassi-Rad

Judging Instinct Exploitation in Statistical Data Explanations Based on Word Embedding
Kang Zhang, Hiroaki Shinden, Tatsuki Mutsuro and Einoshin Suzuki

How Open Source Machine Learning Shapes AI
Max Langenkamp and Daniel Yue

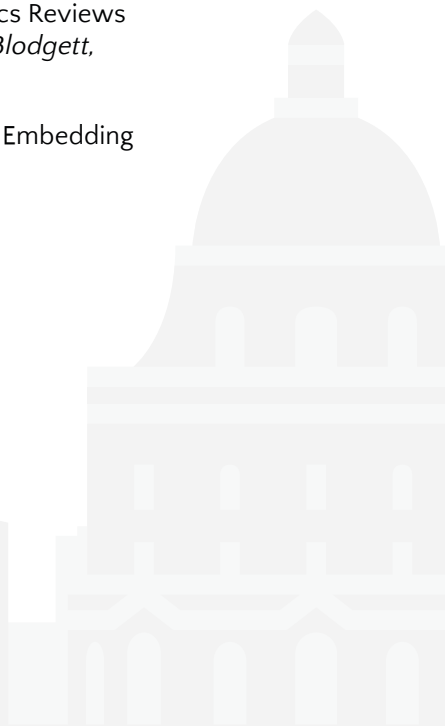
4:00–4:15 **Coffee Break // Arco Building**

4:15–5:00 **Invited Talk 6: Karen Levy, Cornell University // O'Reilly Theatre**
RoboTruckers: The Double Threat of AI for Low-Wage Work

5:00–5:15 **Q&A**

5:15–5:30 **Short Break**

5:30–7:30 **Reception / Poster Session 2 // Arco Building**





9:30–10:45 **Student Lightning Talk Session // O'Reilly Theatre**

Bias in Toxicity Detection
Paula Reyero Lobo

Artificial Intelligence Diffusion in Public Administration
Rohit Madan

Navigating NLP Technologies to Mitigate Harm
Eddie Ungless

Towards Understanding Human-AI Relationships
Marisa Tschopp

Public Attitudes Towards AI
Marina Budic

AI Alignment Dialogues: An Interactive Approach to AI Alignment in Support Agents
Pei-Yu Chen

Emotion Recognition Technology: Re-shaping human relationships
Alexandra Pregent

Hijacking Epistemic Agency - How Emerging Technologies Threaten our Wellbeing as Knowers
John Dorsch

What's (Not) Ideal about Fair Machine Learning?
Otto Sahlgren

Designing objectives and constraints for AI decision-making
Sreejith Balakrishnan

Governing AI Applications to Monitoring and Managing Our Global Environmental Commons
Melissa Chapman

Tell Me Why: Explaining Black-Box Classification Algorithms Using Most Similar Counterfactuals
Julia Barnett

Memorization in Language Models
Katherine Lee

Considering Context as it Relates to Student Agency Conditions in Intelligent Game-Based Learning Environments
Allison Banzon

Fairness Implications behind the Technical Assumptions of Machine Learning
Angelina Wang

Automated Kantian Ethics
Lavanya Singh

Ethics, Transparency, and Accountability for Mass Communication Platforms
Jack Bandy

How to design smart homes for older persons
Nadine Andrea Felber

Comprehensive Human Oversight Framework to ensure accountability over
Autonomous Weapon Systems
Ilse Verdiesen

Understanding algorithmic bias in British English automatic speech recognition
Nina Markl

Rethinking Counterfactual Visual Explanations
Eoin Delaney

Exploitation and Algorithmic Pricing
Arianna Dini

Constructing meaningful Explanations: Logic-based Approaches
Laura State

Dialogue Explanation With Reasoning For AI
Yifan Xu

Examining Trust in Conversational Systems: Conceptual and Empirical Foundations
for the Relation Between User Trust, Related Behavior, and System Trustworthiness
Anuschka Schmitt

Bias in Artificial Intelligence Models in Financial Services
Ángel Pavón Pérez

Ethical Recommenders in the Public Library sector
Savvina Daniil

Socially-Aware Artificial Intelligence for Fair Mobility
Dimitris Michailidis

Artificial Quasi Moral Agency
Jen Semler

WEDNESDAY, AUGUST 3

Ethical Design for AI in Medicine
Éric Pardoux

AI and Legal Personhood: An African Perspective
Meshandren Naidoo

10:45–11:45 **Student Poster Session // Arco Building**

11:45–12:00 **Closing Remarks // O'Reilly Theatre**

12:00–1:00 **Farewell Lunch // Dining Hall**

AIES-22 SPONSORS

AIES 2022 would like to thank the generous sponsors who allowed us to support Ph.D. students, invited speakers, social events, and to reduce the registration fee.

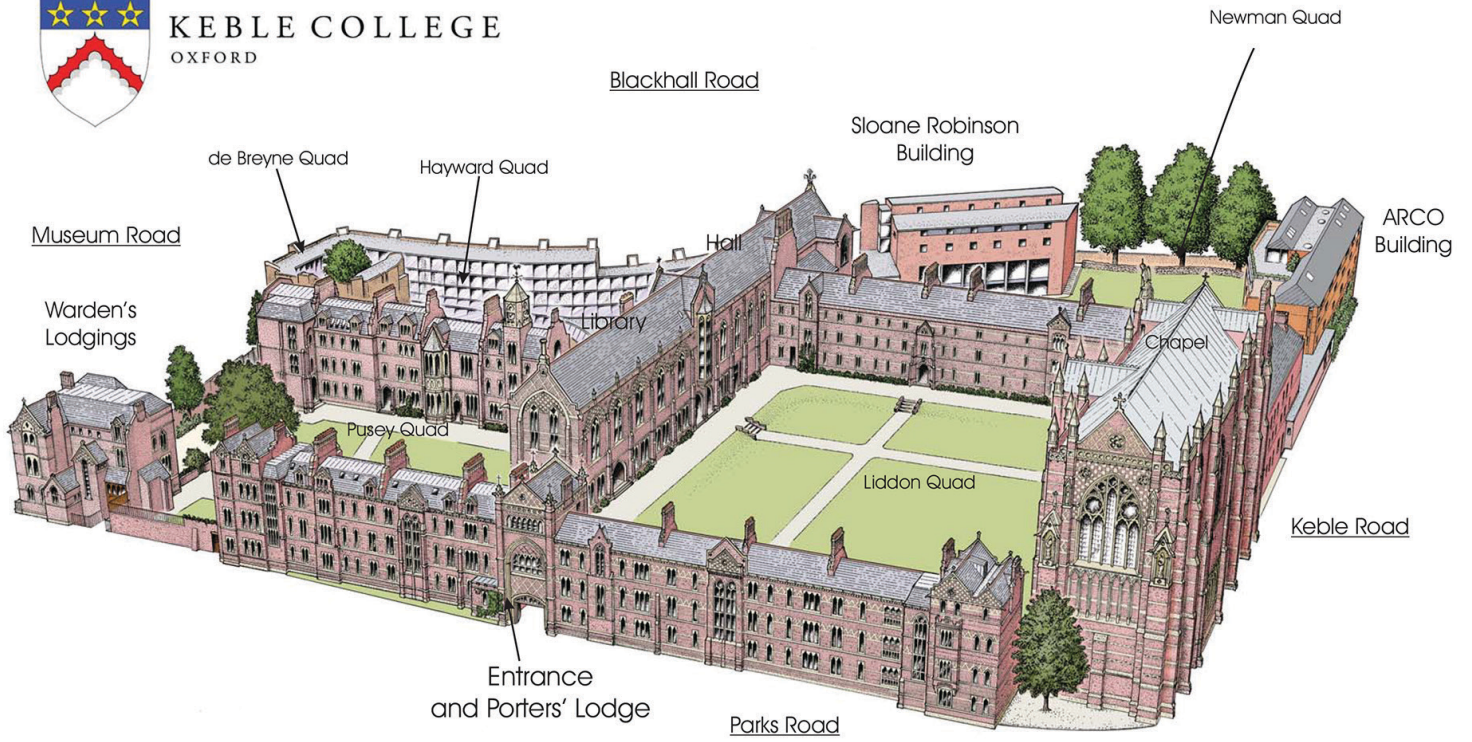


ORGANIZING ASSOCIATIONS





KEBLE COLLEGE
OXFORD



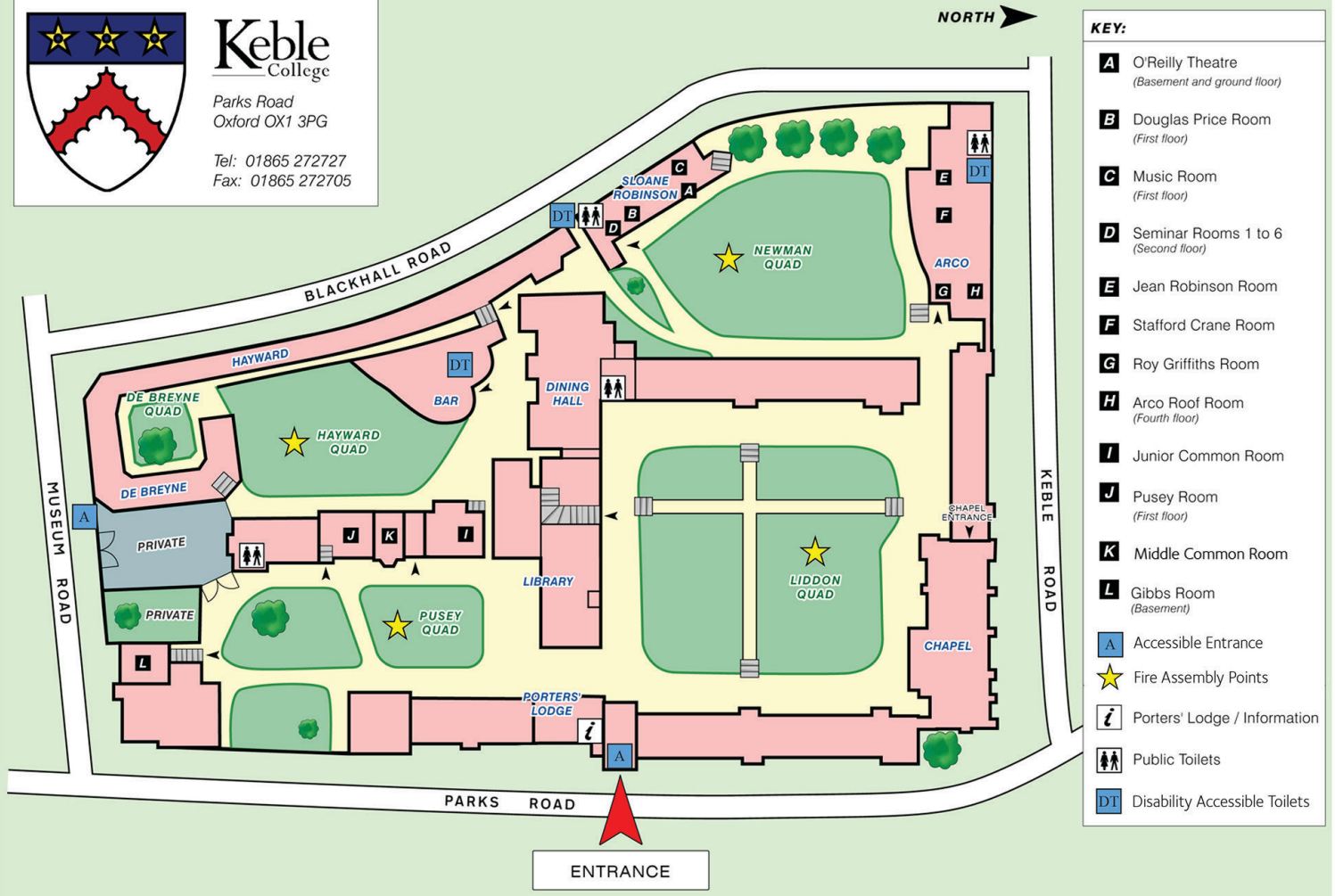
Drawn by Jeremy Bayes, Art-Work-Shop © Keble College 2004.



Keble
College

Parks Road
Oxford OX1 3PG

Tel: 01865 272727
Fax: 01865 272705



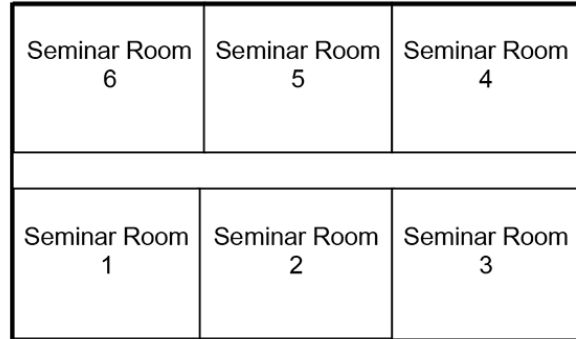
SLOANE ROBINSON BUILDING

Newman Quad - Keble College



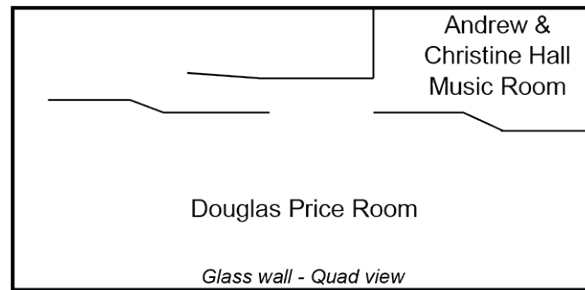
Suite of six seminar rooms

Each seats up to 40 delegates theatre style.



Douglas Price Room

A large multi purpose room.



The O'Reilly Theatre

Seats 250. Theatre built on two levels, using retractable seating and moveable staging for maximum flexibility. State-of-the-art audio visual and sound systems.

