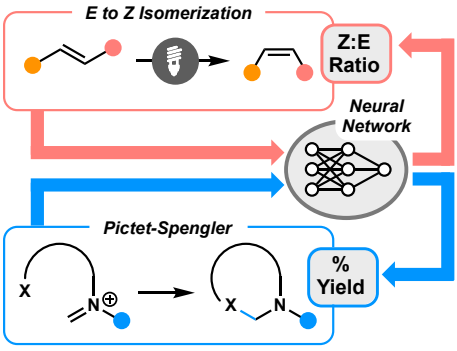
Dr Emma King-Smith emma.king-smith@ed.ac.uk

Research Interests: Organic synthesis, Machine Learning & Artificial Intelligence, AI-guided synthesis

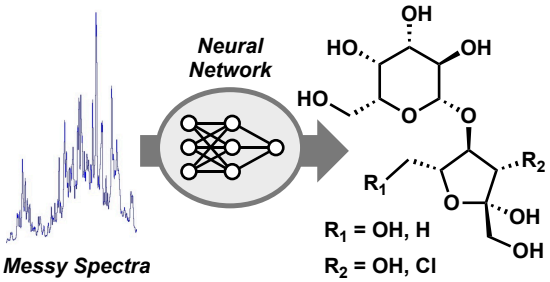
Research projects are centered around bringing machine learning prediction to synthetic organic chemistry. Students will gain experience in coding, machine learning, deep learning, and modern synthetic chemistry. Each project will be a mixture of computational and experimental chemical research. Interested students should contact Emma (emma.king-smith@ed.ac.uk) for an informal discussion.

Project 1: Exploring Experimental Data Requirements for Reaction Prediction: How much data is necessary to achieve accurate machine learning models? Data is the "life blood" of any deep learning model, but its experimental acquisition can be laborious, and time prohibitive. Thus, a keen understanding of the limitations of lower-data environments is crucial for predictive chemical modelling. This project will explore the ideal size and "information entropy" required to model the reaction outcomes of photochemical *E*-alkene to *Z*-alkene isomerizations and Pictet-Spengler reactions, highly useful reactions in the chemists' toolkit.



Project 2: Chemical Knowledge Infusion into Deep Learning Models: Chemistry is a multimodal field: we learn through the words in our textbooks, the descriptors from our lecturers, and the structures in our figures. Despite teaching this way for decades, we have abandoned this method when we teach machine learning models chemistry. Large chemistry models almost always focus solely on the chemistry text *or* chemical structures, not both. This project will investigate methods of infusing chemical knowledge from both visual and language cues into a deep learning model for reaction yield prediction of Buchwald-Hartwig couplings. Prospective reactions will be validated experimentally.

Project 3: Spectra to Structure Elucidation of Compound Mixtures with Machine Learning: ^{1}H NMR is a workhorse structure elucidation method in organic chemistry. Naturally, it is incredibly important to know what compound one has made in both fundamental and applied research, thus accurate structural assignment is indispensable. Whilst many ML and DFT-derived methods have been developed to predict spectra from structure the reverse is much more challenging. To date, no method has ever been able to discern mixtures of compounds. This project will investigate how convolutional neural networks can be used to elucidate the structures of organic compounds from their ^{1}H NMR spectra, including mixtures of compounds.



Recent Publications:

1. "Deconvoluting Low Yield from Weak Potency in Direct-to-Biology Workflows with Machine Learning", *RSC Med. Chem.* **2024**, *15*, 1015–1021.
2. "Transfer Learning for a Foundational Chemistry Model", *Chem. Sci.* **2024**, *15*, 5143–5151.
3. "Predictive Minisci Late Stage Functionalization with Transfer Learning", *Nature Commun.* **2024**, *15*, 426–438.
4. "Probing the Chemical 'Reactome' with High Throughput Experimentation", *Nature Chem.* **2024**, *16*, 633–643.