FONDAZIONE
EDMUND MACH
dal 1874

SEMINAR

# AI -enabled scientific discovery in image collections

**09 July 2025**
**h. 14:30 - 15:30**

**Sala Gialla CTT**
**Fondazione Edmund Mach**
**SAN MICHELE A/ADIGE**

**www.fmach.it**


**Dr. Sara Beery**
Assistant Professor
MIT EECS' Faculty of AI and Decision Making and CSAIL, USA



Natural world images collected by communities of enthusiast volunteers provide a vast and largely uncurated source of data. For instance, iNaturalist has over 200 million images tagged with species labels, already contributing immensely to research such as biodiversity monitoring and having been cited in over 4,000 scientific papers. Yet, these images are also known to contain a wealth of "secondary data" captured unintentionally or otherwise included in images and not properly reflected in image labels. Although this data contains crucial insights into interactions, animal social behavior, morphology, habitat, co-occurrence, and many more questions, the costly, time-consuming, or expert-dependent analysis needed to extract such information prevents breakthroughs. Advances in deep learning methods for language and computer vision have the potential to enable the efficient and automated processing techniques needed to unlock the "hidden treasure" in such datasets– being able to directly search large image collections for these concepts would enable richer analyses that span beyond species identification. We propose interactive, open-ended image retrieval as a mechanism to support scientific discovery in these collections, and introduce INQUIRE, a novel text-to-image retrieval benchmark built to provide a rigorous evaluation that challenges models to demonstrate advanced knowledge and visual reasoning on expert, scientifically impactful retrieval tasks.