



THESE KINDS OF MACHINE LEARNING MODELS ARE OFTEN-TRAINED ON A 2012 DATASET CALLED **IMAGENET**.

IMAGENET CONTAINS **14,197,122** IMAGES, SCRAPED FROM **THE INTERNET**.

THESE IMAGES WERE INDIVIDUALLY TAGGED BY CONTENT WORKERS ALL AROUND THE WORLD (WHO WERE OFTEN MAKING FAR LESS THAN MINIMUM WAGE).

WORKERS COULD CHOOSE TO TAG IMAGES FROM 22,000 'CLASSES', BORROWED FROM A LEXICAL DATABASE CALLED **WORDNET**.

THE CLASSES THEMSELVES CAME FROM GRAD STUDENTS WHO, IN THE 1990S, HAND-TAGGED TEXT CONTENT FROM **THE WALL STREET JOURNAL**.

A MODEL TRAINED ON IMAGENET THEN, KNOWS ONLY WHAT IT HAS LEARNED FROM THE **INTERNET** (MEMES AND ALL), READ THROUGH THE LENS OF A NEWSPAPER FROM **NEW YORK CITY**, WRITTEN FOR MIDDLE-AGED STOCK TRADERS.

IT IS A MODEL THAT KNOWS A LOT ABOUT CATS...

AND VERY LITTLE ABOUT PUERTO RICO IN THE 1940S.