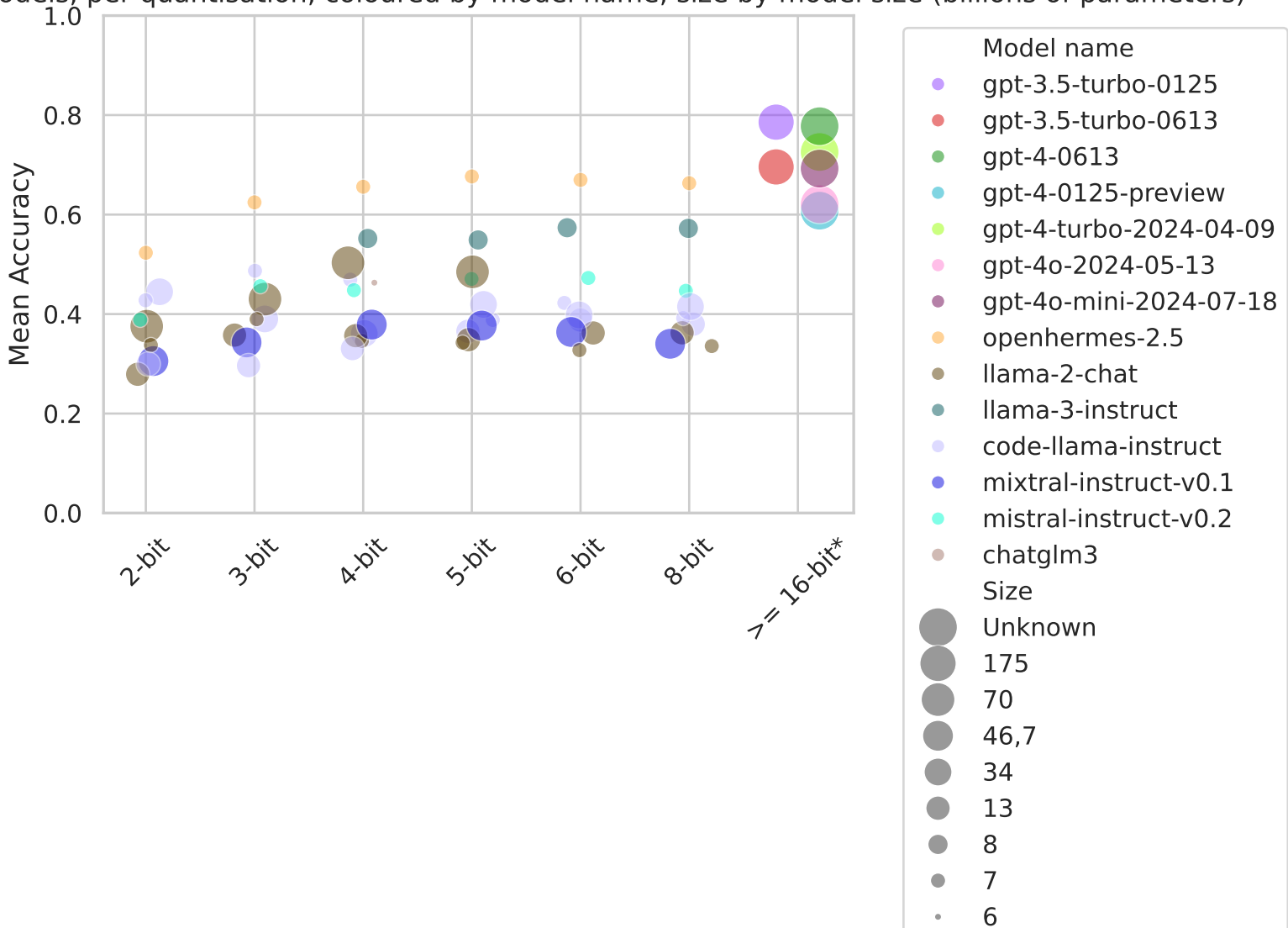Scatter plot across models, per quantisation, coloured by model name, size by model size (billions of parameters)