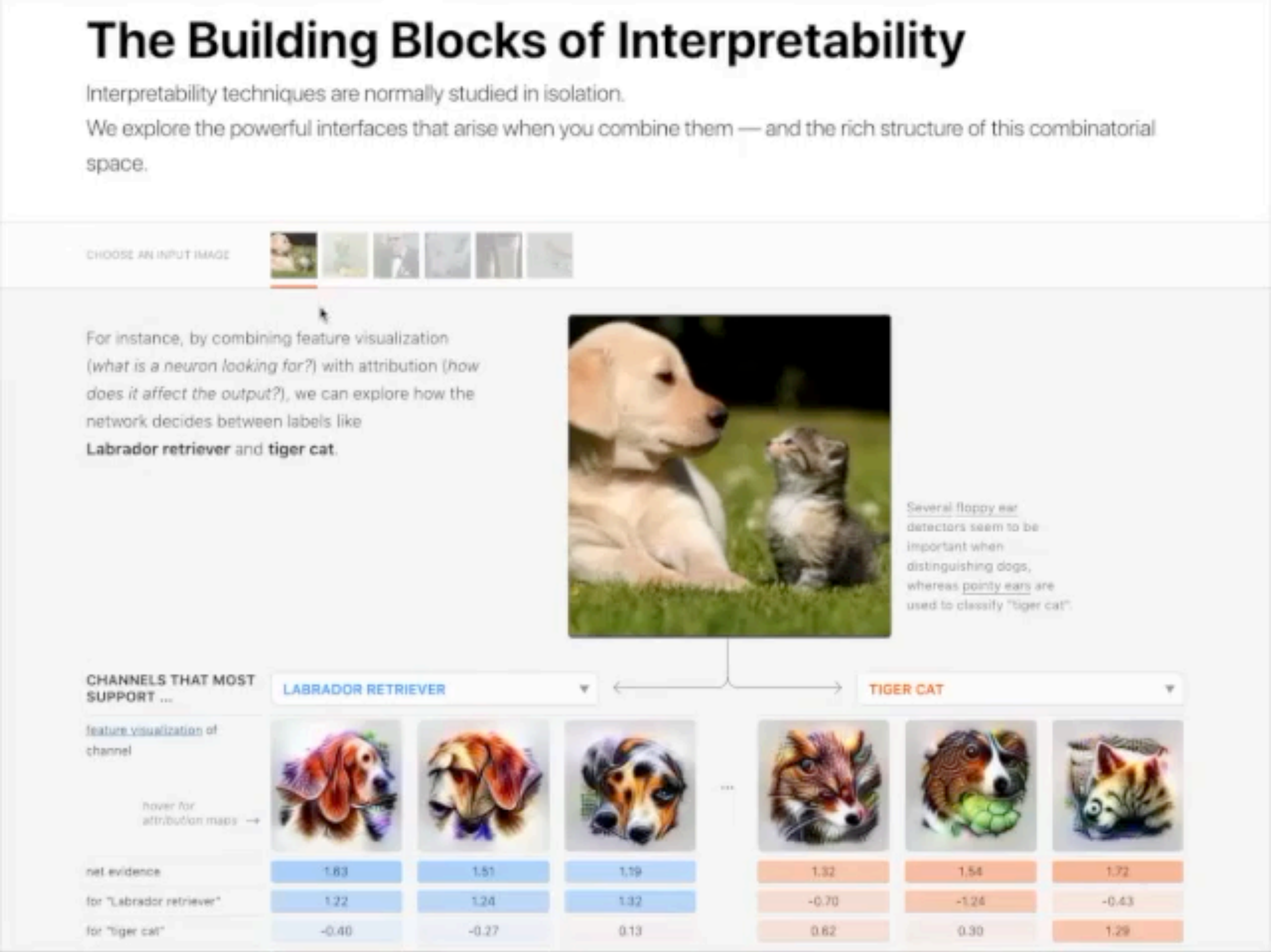
exploratory explanatory

Interaction + Documents



The Building Blocks of Interpretability

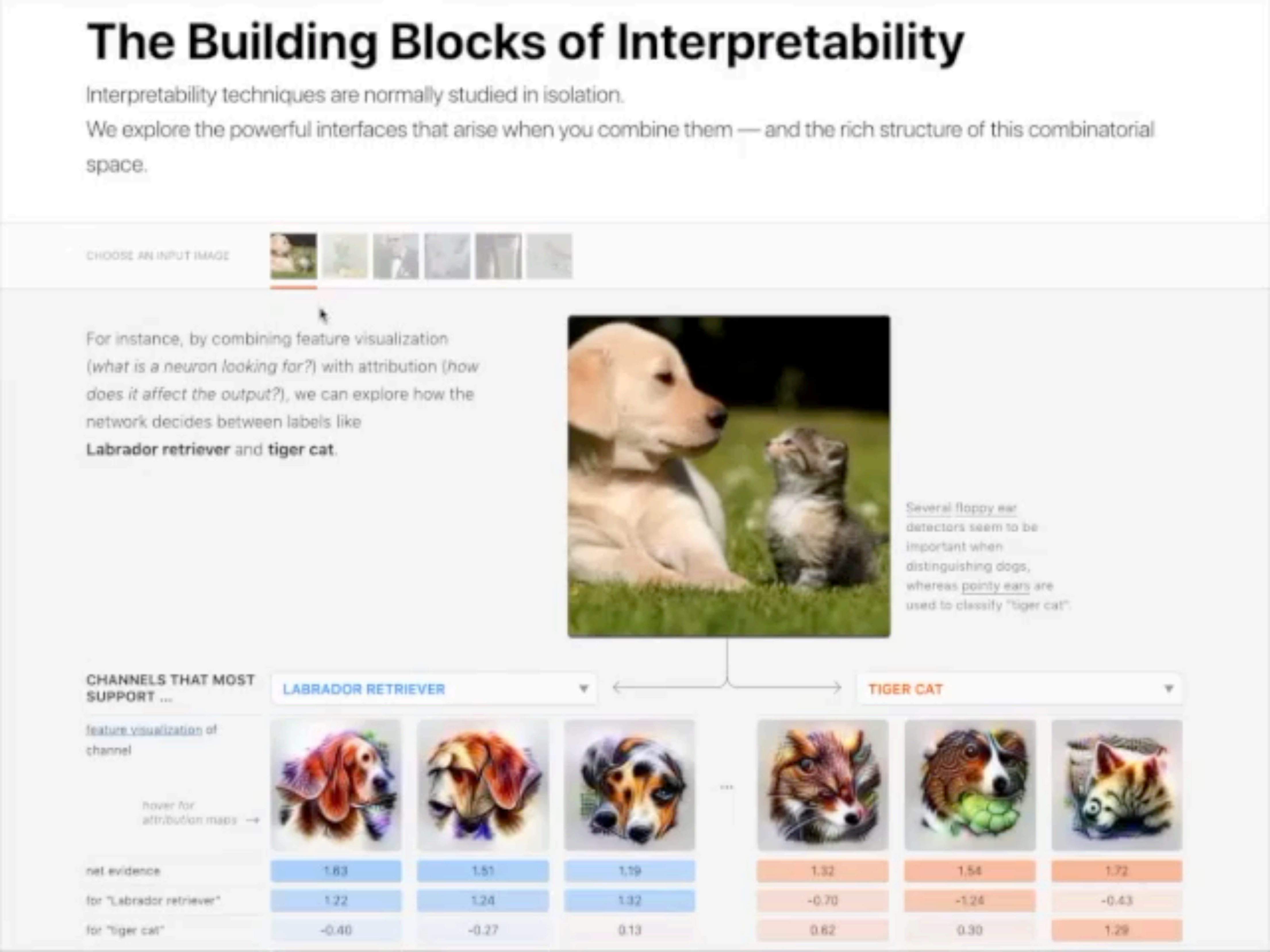
Interpretability techniques are normally studied in isolation.

We explore the powerful interfaces that arise when you combine them — and the rich structure of this combinatorial space.

For instance, by combining feature visualization (*what is a neuron looking for?*) with attribution (*how does it affect the output?*), we can explore how the network decides between labels like **Labrador retriever** and **tiger cat**.

Several floppy ear detectors seem to be important when distinguishing dogs, whereas pointy ears are used to classify "tiger cat".

CHANNELS THAT MOST SUPPORT ...	LABRADOR RETRIEVER ▾				TIGER CAT ▾		
feature visualization of channel hover for attribution maps →				...			
net evidence	1.83	1.51	1.19		1.32	1.54	1.72
for "Labrador retriever"	1.22	1.24	1.32		-0.70	-1.24	-0.43
for "tiger cat"	-0.40	-0.27	0.13		0.62	0.30	1.29



The Building Blocks of Interpretability

Interpretability techniques are normally studied in isolation.
We explore the powerful interfaces that arise when you combine them — and the rich structure of this combinatorial space.

For instance, by combining feature visualization (*what is a neuron looking for?*) with attribution (*how does it affect the output?*), we can explore how the network decides between labels like **Labrador retriever** and **tiger cat**.

Several floppy ear detectors seem to be important when distinguishing dogs, whereas pointy ears are used to classify "tiger cat".

CHANNELS THAT MOST SUPPORT ...	LABRADOR RETRIEVER ▾				TIGER CAT ▾		
feature visualization of channel hover for attribution maps →				...			
net evidence	1.83	1.51	1.19		1.32	1.54	1.72
for "Labrador retriever"	1.22	1.24	1.32		-0.70	-1.24	-0.43
for "tiger cat"	-0.40	-0.27	0.13		0.62	0.30	1.29

Future of Visualization Tools

Design for engaging **exploratory** and **explanatory** data-driven stories

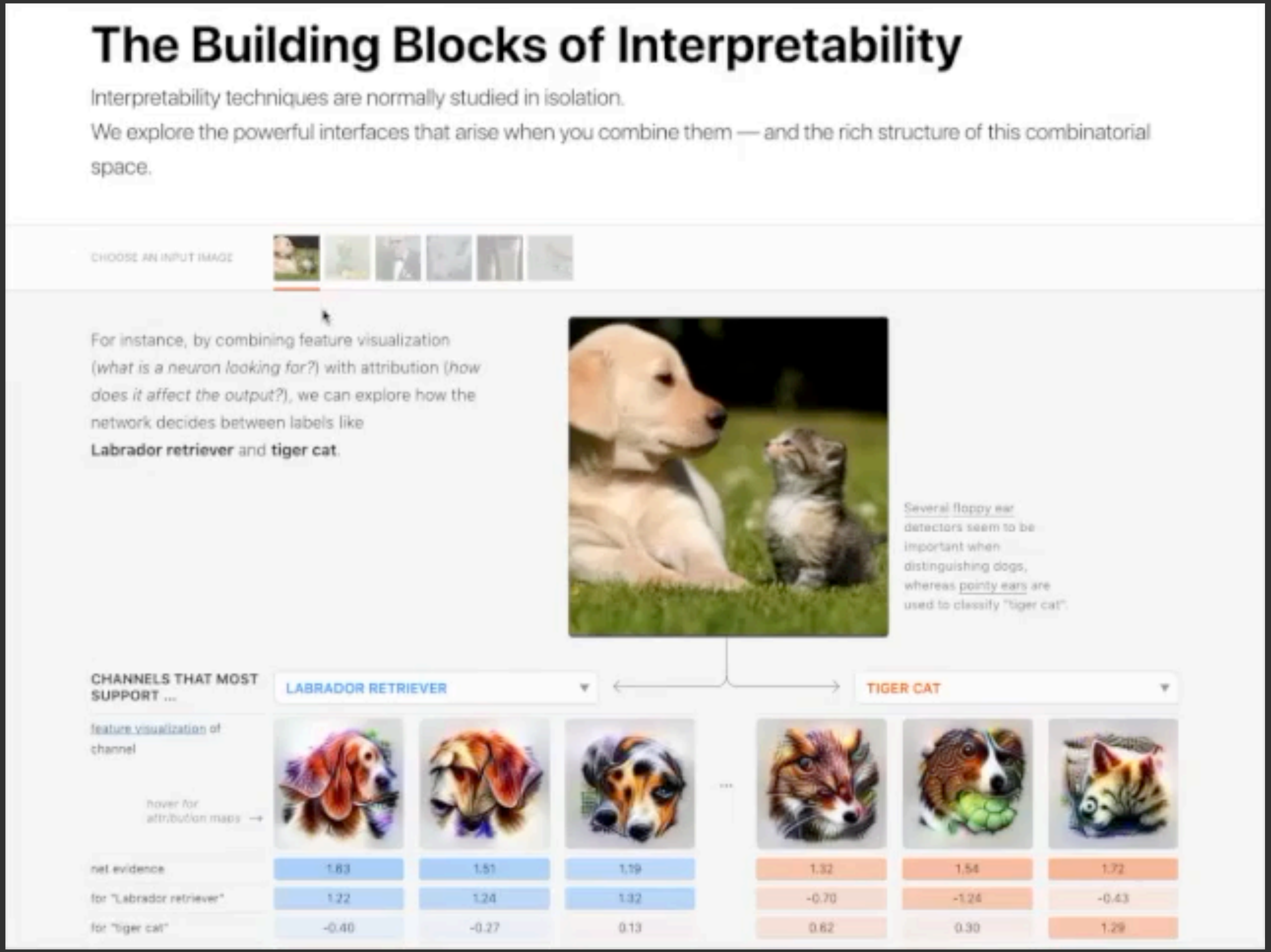
Interaction + Videos

Interaction + Comics

Interaction + Documents

Interaction + Spreadsheets

and more…



The Building Blocks of Interpretability

Interpretability techniques are normally studied in isolation.

We explore the powerful interfaces that arise when you combine them — and the rich structure of this combinatorial space.

For instance, by combining feature visualization (*what is a neuron looking for?*) with attribution (*how does it affect the output?*), we can explore how the network decides between labels like **Labrador retriever** and **tiger cat**.

Several floppy ear detectors seem to be important when distinguishing dogs, whereas pointy ears are used to classify "tiger cat".

CHANNELS THAT MOST SUPPORT …	LABRADOR RETRIEVER			…	TIGER CAT		
feature visualization of channel hover for attribution maps →							
net evidence	1.63	1.51	1.19		1.32	1.54	1.72
for "Labrador retriever"	1.22	1.24	1.32		-0.70	-1.24	-0.43
for "tiger cat"	-0.40	-0.27	0.13		0.62	0.30	1.29

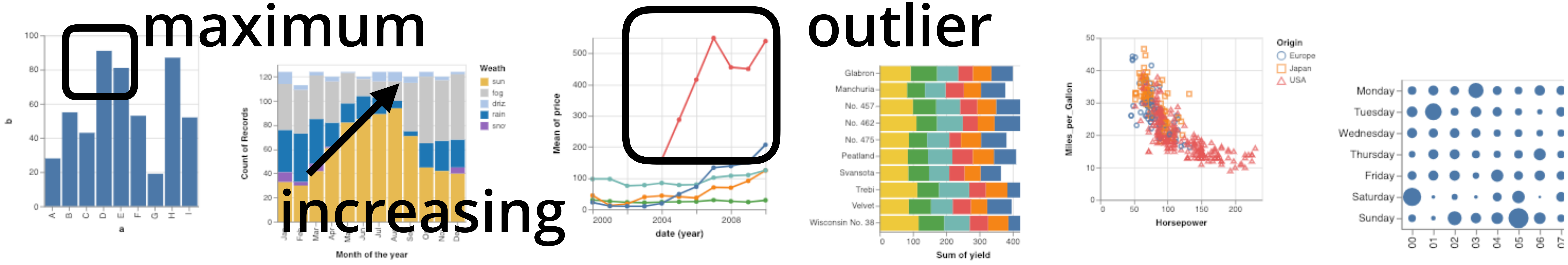
Distill'18

Future of Visualization Tools

Design for engaging

exploratory and **explanatory** data-driven stories

How to provide automatic assistance for discovering story ideas?


maximum
increasing
outlier
b
a
Count of Records
Month of the year
Weath
sun
fog
driz
rain
sno
Mean of price
date (year)
Sum of yield
Glabron
Manchuria
No. 457
No. 462
No. 475
Peatland
Svansota
Trebi
Velvet
Wisconsin No. 38
Miles_per_Gallon
Horsepower
Origin
Europe
Japan
USA
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday