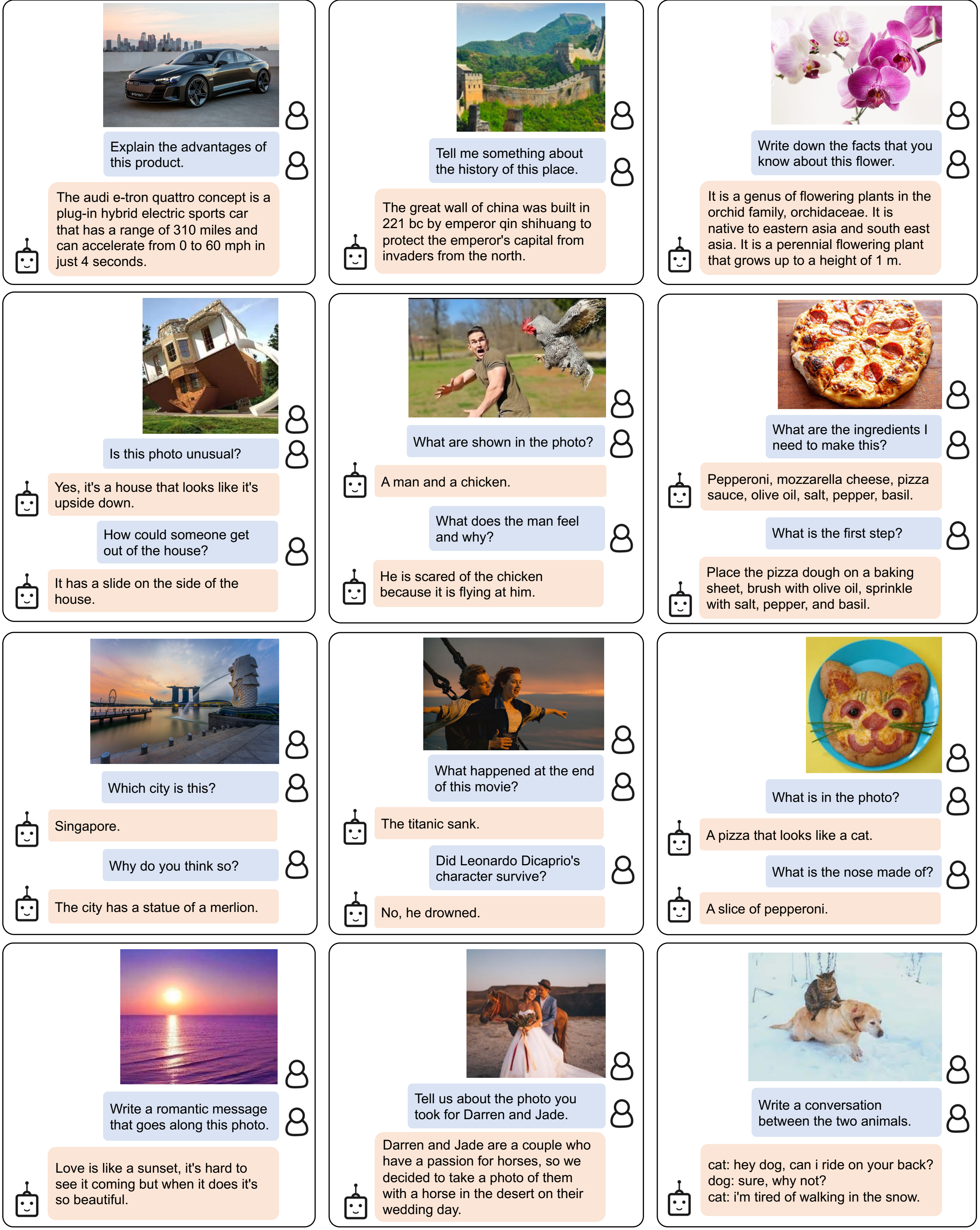

Explain the advantages of this product.

The audi e-tron quattro concept is a plug-in hybrid electric sports car that has a range of 310 miles and can accelerate from 0 to 60 mph in just 4 seconds.

Tell me something about the history of this place.

The great wall of china was built in 221 bc by emperor qin shihuang to protect the emperor's capital from invaders from the north.

Write down the facts that you know about this flower.

It is a genus of flowering plants in the orchid family, orchidaceae. It is native to eastern asia and south east asia. It is a perennial flowering plant that grows up to a height of 1 m.

Is this photo unusual?

Yes, it's a house that looks like it's upside down.

How could someone get out of the house?

It has a slide on the side of the house.

What are shown in the photo?

A man and a chicken.

What does the man feel and why?

He is scared of the chicken because it is flying at him.

What are the ingredients I need to make this?

Pepperoni, mozzarella cheese, pizza sauce, olive oil, salt, pepper, basil.

What is the first step?

Place the pizza dough on a baking sheet, brush with olive oil, sprinkle with salt, pepper, and basil.

Which city is this?

Singapore.

Why do you think so?

The city has a statue of a merlion.

What happened at the end of this movie?

The titanic sank.

Did Leonardo Dicaprio's character survive?

No, he drowned.

What is in the photo?

A pizza that looks like a cat.

What is the nose made of?

A slice of pepperoni.

Write a romantic message that goes along this photo.

Love is like a sunset, it's hard to see it coming but when it does it's so beautiful.

Tell us about the photo you took for Darren and Jade.

Darren and Jade are a couple who have a passion for horses, so we decided to take a photo of them with a horse in the desert on their wedding day.

Write a conversation between the two animals.

cat: hey dog, can i ride on your back?
dog: sure, why not?
cat: i'm tired of walking in the snow.

Figure 4. Selected examples of **instructed zero-shot image-to-text generation** using a BLIP-2 model w/ ViT-g and FlanT5$_{\text{XXL}}$, where it