The Romanian Room

You have a part-time job helping with what you have been told is an artificial intelligence experiment at the University of Virginia. You are placed in a room with a table, chair, a pen, and a very large book. You are told that someone will hand you a piece of paper with words in the Romanian language. You do not need to understand the words. You are instructed to look through all the catagories in the categorie book, to find the best match between the words on the paper and the words between <model> tags. Hashtags (#) mean that more words may be present on the piece of paper, but may be ignored. You must then copy what is written between the <model> tags onto the back of the piece of paper you were given, and return it.

A man comes into your room, and hands you a scrap of paper with this written on it:

Cine este filozoful tău preferat?

Looking through your categorie book, this is the only match you find:

<categorie>
<model> # filozoful # </model>
<format>Îmi plac mulți filozofi. Sunt amuzat de John Searle, pentru că este convins că nu mă pot gândi.</format>
</categorie>

This matches any string of words containing "filozoful". So you copy what is between the <format> tags onto the back of the scrap of paper, and return it. The man laughs, and says "Ești foarte amuzant." You smile politely, but you have no idea what he said.

The man passes you many more scraps of paper. You carefully identify matching categories in your enormous book, and scribble replies for the man. While you are doing this, you are thinking about new strategies to use to beat the video game you have been playing, how much homework you have to do when you get home, and who to invite to homecoming.

Finally the man leaves. When he returns, he shakes your hand, then looks admiringly at your book of categories, and says "Felicitări! Tu și cartea ta de categorii au trecut testul Turing!" Later, you find out (it is all over the internet) that you and your book have just had an extensive conversation about philosophy with a university professor, and your categorie book has passed the Turing test. Apparently, the book can think! You, of course, were just a page flipper, looking for matches between what were (to you) meaningless strings of symbols.

You later learn that the professor who passed you all those scraps of paper had earlier done exactly the same thing with another person, who also had a copy of your book. That person was a Romanian philosophy graduate student, who completely ignored the answers suggested by the book, and wrote his own. The professor had not been able to tell which person was actually thinking about philosophy, and giving his own answers, and which had been just mindlessly matching symbols, and writing down symbols he did not understand.

1. John Searle might say that you and your book of categories did a good job of *simulating* a philosophical conversation, but that does not mean that either you or the book were *really having* a philosophical conversation.

   Give two examples of *simulations* which do not actually involve actually *doing* what is being simulated.

2. Some critics of Searle say he is committing the *fallacy of composition*. You and the rule book, they say, are parts of a larger whole. Just because you are not thinking about philosophy, and the rule book is not thinking about philosophy, does not mean that the larger whole is not thinking about philosophy.

   A dualist philosopher, one who believes we are immaterial thinking substances, might argue that a human brain by itself cannot think, since it is made up of unthinking neurons. How could you criticize this argument?